(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 September 2006 (08.09.2006)

PCT

(10) International Publication Number
WO 2006/094238 A2

(51) International Patent Classification:
C12N 15/11 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/US2006/007747

(22) International Filing Date: 3 March 2006 (03.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/658,248	3 March 2005 (03.03.2005)	US
60/705,631	3 August 2005 (03.08.2005)	US
60/732,539	1 November 2005 (01.11.2005)	US

(71) Applicant *(for all designated States except US)*: **ISIS PHARMACEUTICALS, INC.** [US/US]; 1896 Rutherford Rd, Carlsbad, CA 92008 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **SAMPATH, Rangarajan** [US/US]; 12223 Mannix Road, San Diego, CA 92129 (US).

(74) Agents: **CASIMIR, David** et al.; Medlen and Carroll, LLP, Suite 305, 101 Howard Street, San Francisco, California 94105 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS FOR USE IN IDENTIFICATION OF ADVENTITIOUS VIRUSES

(57) Abstract: The present invention provides compositions, kits and methods for rapid identification and quantification of adventitious contaminant viruses by molecular mass and base composition analysis.

WO 2006/094238 A2

COMPOSITIONS FOR USE IN IDENTIFICATION OF ADVENTITIOUS VIRUSES

RELATED APPLICATIONS

[01] The present application 1) claims the benefit of priority to U.S. Provisional Application Serial No. 60/658,248, filed March 3, 2005; 2) claims the benefit of priority to U.S. Provisional Application Serial No. 60/705,631, filed August 3, 2005; 3) claims the benefit of priority to U.S. Provisional Application Serial No. 60/732,539, filed November 1, 2005 and 4) claims the benefit of priority to U.S. Provisional Application Serial No. 60/740,617, filed November 28, 2005. Each of the above listed U.S. Provisional Applications is incorporated herein by reference in entirety. Methods disclosed in U.S. Application Serial Nos. 10/156,608, 09/891,793, 10/418,514, 10/660,997, 10/660,122, 10,660,996, 10/660,998, 10/728,486, 10/405,756, 11/060,135, and 11/073,362, are commonly owned and incorporated herein by reference in their entirety for any purpose.

STATEMENT OF GOVERNMENT SUPPORT

[02] This invention was made with United States Government support under NIH contract N01 AI40100. The United States Government may have certain rights in the invention.

FIELD OF THE INVENTION

[03] The present invention provides compositions, kits and methods for rapid identification and quantification of adventitious contaminant viruses by molecular mass and base composition analysis.

BACKGROUND OF THE INVENTION

A. Adventitious Viruses

[04] Adventitious viruses represent a major risk associated with the use of cell-substrate derived biologicals, including vaccines and antibodies, for human use. The possibility for viral contamination exists in primary cultures and established cultures, as well as Master Cell Banks, end-of-production cells, and bulk harvest fluids. This is a major obstacle to the use of neoplastic-immortalized cells for which the mechanism of transformation is unknown is that these could have a higher risk of containing oncogenic viruses. Extensive testing for the presence of potential extraneous agents is therefore required to ensure the safety of the vaccines. Among the methods used for this purpose are animal inoculations, electron microscopy and in vitro molecular and antibody assays that provide a screen for viral agents. Another

critical consideration for assessing the safety concerns associated with viral vaccines is the detection of endogenous retroviral sequences while using avian, murine, non-human primate, and human cell lines. Endogenous retroviral sequences are an integral part of eukaryotic genomes, and while the majority of these sequences are defective, a few can produce infectious virus, either spontaneously upon long-term culture. These can also be induced upon treatment with various chemical or other agents that may be part of the normal production system. The activation of an endogenous, infectious retrovirus in a cell substrate that is used for the production of biologics is an important safety concern, especially in the case of live, viral vaccines, where minimal purification and inactivation steps are used in order to preserve high vaccine potency.

[05] The currently established methods for measuring RT-activity include the highly sensitive, product-enhanced reverse transcriptase assays (PERT) that can detect 1–10 virions and transmission electron microscopy (TEM) to analyze infective retroviruses particles. However, the above techniques are not specific and do not provide any information regarding the source of the RT activity. PCR-based detection of retroviruses can be used in combination with other assays such as reverse transcriptase, electron microscopy infectivity or co-cultivation to increase the sensitivity of detection or to identify a particular adventitious agent present in the test sample. Further, while some studies demonstrate that a low level of RT activity is not generally associated with a replicating agent; major concerns remain regarding the consequences of the presence of such non-productive, non-replicating defective infections in the vaccine, as there is the potential for integration into the host genome.

[06] Retrovirus-induced tumorigenesis can involve the generation of a novel pathogenic virus by recombination between replication-competent and -defective sequences and/or activation of a cellular oncogene by a long terminal repeat (LTR) due to upstream or downstream insertion of retrovirus sequences. To address the possible integration of extraneous retroviral sequences in human cells by RT-containing particles, multiple PCR strategies have been used. These include direct PCR of DNase-treated inoculum using primers from the highly conserved pol region and Alu PCR using LTR primers in conjunction with Alu primers that specifically amplify viral-cellular DNA junctions of integrants.

[07] Future strategies to detect adventitious agents must address three fundamental problems. First, there are large numbers of known viral agents that are potential contaminants, each with a large number of potential strain variants. Second, history has shown that not all adventitious agents fall into anticipated families of viruses, so unanticipated virus families must also be considered. Third, the test must be practical to perform on a large number of samples in a standardized, high-throughput, quality-controlled fashion. The premise of this proposal is that we can leverage recently developed and validated methods using mass spectrometry analysis of broad-range PCR reactions for rapid, sensitive, cost-effective

detection of broad ranges of adventitious agents, including previously unknown/uncharacterized viruses and endogenous retroviruses.

B. Drug Resistance

[08] Drug resistance in bacteria and viruses is frequently mediated by point mutations in key genes whose gene products interact directly or indirectly with the drug. While there are several methods available for identification of single nucleotide polymorphisms (SNPs) in nucleic acid sequences, the functional unit that encodes each amino acid is the codon, where three successive nucleotides are responsible for encoding each amino acid. Mutations in any of the three nucleotides may or may not result in a mutation in the encoded amino acid, depending upon the particular amino acid and the rules of the genetic code. Because the genetic code is deciphered as a sequence, both the identity and the order of the nucleotides are important in determining the encoded amino acid. Thus, DNA sequencing has become the method of choice for analysis of mutations that result in amino acid changes. DNA sequencing has significant disadvantages as an analysis method for routine use a clinical laboratory setting. It is still relatively expensive and labor intensive, and thus is used only for very important analyses. An example of this is determination of drug resistance in viruses such as HIV and in bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA). Drug resistance in HIV has now emerged as a significant problem in both untreated and drug-treated patient populations. The decision to select a particular drug-treatment regimen that the virus will respond to is critical to success of therapy. Drug resistance testing has been shown to improve the clinical outcome in HIV-infected individuals and thus is now recommended for new infections or for patients infected as long as two years or more prior to initiating therapy, in the case of antiretroviral failures and during pregnancy. Thus, despite the costs, DNA sequencing is currently being used for determination of viral drug resistance. Typically, a serum sample is analyzed by PCR amplification of the reverse transcriptase and protease genes, followed by sequencing of approximately 900 nucleotides of the reverse transcriptase gene and 300 nucleotides of the protease gene. The DNA sequence is then used to determine the optimal drug regimen. A drawback of sequencing is that DNA sequencing technology for identification of drug-resistant viruses is that it is not easily able to identify the components present in a mixed sample, particularly in a scenario where a fraction of the virus population has mutated. DNA sequencing was developed on the assumption that the sample being analyzed is homogeneous. However, the HIV populations that infect humans are not homogeneous, and RNA viruses such as HIV are known to rapidly mutate, creating a population of mixed sequences in each infected individual. In the presence of drug selection, mutations that mediate drug resistance that occur at low frequency grow with a selective advantage and eventually can dominate the population, causing treatment failure. In this scenario, the mutant virus starts out as an undetectable fraction of the population which increases to a higher percentage over time. It would be valuable to

identify drug resistant virus populations early, before they have a chance to increase the viral load. DNA sequencing methods can identify mixed populations, but do so poorly. In a recent publication using the ABI PRISM 3100 genetic analyzer, it was reported that a viral mixture containing approximately 40% of the mutant viral population can be detected with 95% confidence. However, 40% of a typical viral load (1,800 to 10,500 HIV copies/ml) means a blood burden (assuming 5 liters of blood) of up to 21 million drug-resistant viral copies. Other analytical methods are capable of identifying mutations with more sensitivity than sequencing, but these methods are time consuming, laborious and not amenable to high throughput processes.

[09] Thus, there is a need for rapid and cost effective methods that can be applied as alternatives to sequencing in genomic analysis for variations that mediate amino acid changes. The present invention satisfies this need. The present invention provides, *inter alia*, methods of identifying adventitious contaminant viruses. Also provided are oligonucleotide primers, compositions and kits containing the oligonucleotide primers, which produce amplification products whose molecular masses provide the means to identify adventitious contaminant viruses at the sub-species level.

SUMMARY OF THE INVENTION

[10] The present invention provides compositions, kits and methods for rapid identification and quantification of adventitious contaminant viruses by molecular mass and base composition analysis.

[11] One embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 47.

[12] Another embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 286.

[13] Another embodiment is a composition of is an oligonucleotide primer pair including an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 47 and an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 286.

[14] One embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 70.

[15] Another embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 286.

[16] Another embodiment is a composition of is an oligonucleotide primer pair including an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 70 and an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 286.

[17] One embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 165.

[18] Another embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 286.

[19] Another embodiment is a composition of is an oligonucleotide primer pair including an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 165 and an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 286.

[20] One embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 122.

[21] Another embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 275.

[22] Another embodiment is a composition of is an oligonucleotide primer pair including an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 122 and an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 275.

[23] One embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 100.

[24] Another embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 336.

[25] Another embodiment is a composition of is an oligonucleotide primer pair including an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID

NO: 100 and an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 336.

[26] One embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 61.

[27] Another embodiment is an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 324.

[28] Another embodiment is a composition of is an oligonucleotide primer pair including an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 61 and an oligonucleotide primer 14 to 35 nucleobases in length having at least 70% sequence identity with SEQ ID NO: 324.

[29] In some embodiments, either or both of the primers of the primer pair contain at least one modified nucleobase such as 5-propynyluracil or 5-propynylcytosine for example.

[30] In some embodiments, either or both of the primers of the primer pair comprises at least one universal nucleobase such as inosine for example.

[31] In some embodiments, either or both of the primers of the primer pair comprises at least one non-templated T residue on the 5'-end.

[32] In some embodiments, either or both of the primers of the primer pair comprises at least one non-template tag.

[33] In some embodiments, either or both of the primers of the primer pair comprises at least one molecular mass modifying tag.

[34] Some embodiments are kits that contain the primer pair compositions. In some embodiments, each member of the one or more primer pairs of the kit is of a length of 14 to 35 nucleobases and has 70% to 100% sequence identity with the corresponding member from the group of primer pairs represented by SEQ ID NOs: 70:286, 165:286, 122:275, 100:336, and 61:324.

[35] Some embodiments of the kits contain at least one calibration polynucleotide.

[36] Some embodiments of the kits contain at least one anion exchange functional group linked to a magnetic bead.

[37] In some embodiments, the present invention provides primers and compositions comprising pairs of primers, and kits containing the same, and methods for use in identification of adventitious contaminant viruses. The primers are designed to produce amplification products of DNA encoding genes that have conserved and variable regions across a given viral family. The invention further provides compositions comprising pairs of primers and kits containing the same, which are designed to provide species and sub-species characterization of adventitious contaminant viruses.

[38] In some embodiments, the present invention also provides methods for identification of adventitious contaminant viruses. Nucleic acid from the virus is amplified using the primers described above to obtain an amplification product. The molecular mass of the amplification product is measured. Optionally, the base composition of the amplification product is determined from the molecular mass. The molecular mass or base composition is compared with a plurality of molecular masses or base compositions of known adventitious contaminant virus identifying amplicons, wherein a match between the molecular mass or base composition and a member of the plurality of molecular masses or base compositions identifies the adventitious contaminant virus. In some embodiments, the molecular mass is measured by mass spectrometry.

[39] In some embodiments, the present invention is also directed to a method for determining the presence or absence of an adventitious contaminant virus in a sample. Nucleic acid from the sample is amplified using the composition described above to obtain an amplification product. The molecular mass of the amplification product is determined. Optionally, the base composition of the amplification product is determined from the molecular mass. The molecular mass or base composition of the amplification product is compared with the known molecular masses or base compositions of one or more known adventitious contaminant virus identifying amplicons, wherein a match between the molecular mass or base composition of the amplification product and the molecular mass or base composition of one or more known adventitious contaminant virus identifying amplicons indicates the presence of the adventitious contaminant virus in the sample. In some embodiments, the molecular mass is measured by mass spectrometry.

[40] In some embodiments, the present invention also provides methods for determination of the quantity of an unknown adventitious contaminant virus in a sample. The sample is contacted with the composition described above and a known quantity of a calibration polynucleotide comprising a calibration sequence. Nucleic acid from the unknown adventitious contaminant virus in the sample is

concurrently amplified with the composition described above and nucleic acid from the calibration polynucleotide in the sample is concurrently amplified with the composition described above to obtain a first amplification product comprising an adventitious contaminant virus identifying amplicon and a second amplification product comprising a calibration amplicon. The molecular mass and abundance for the adventitious contaminant virus identifying amplicon and the calibration amplicon is determined. The adventitious contaminant virus identifying amplicon is distinguished from the calibration amplicon based on molecular mass, wherein comparison of adventitious contaminant virus identifying amplicon abundance and calibration amplicon abundance indicates the quantity of adventitious contaminant virus in the sample. In some embodiments, the base composition of the adventitious contaminant virus identifying amplicon is determined.

[41] In some embodiments, the present invention provides methods for detecting or quantifying adventitious contaminant virus by combining a nucleic acid amplification process with a mass determination process. In some embodiments, such methods identify or otherwise analyze the adventitious contaminant virus by comparing mass information from an amplification product with a calibration or control product. Such methods can be carried out in a highly multiplexed and/or parallel manner allowing for the analysis of as many as 300 samples per 24 hours on a single mass measurement platform. The accuracy of the mass determination methods in some embodiments of the present invention permits allows for the ability to discriminate between different adventitious viruses such as members of the following families: p\
Papillomaviridae, Polyomaviridae, Retroviridae, Parvoviridae, Herpesviridae (Human herpesviruses 1 through 8, Bovine herpesvirus, Canine herpesvirus and Simian cytomegalovirus), Hepadnaviridae (Hepatitis B virus), Hepeviridae (Hepatitis E virus), Deltavirus (Hepatitis delta virus), Adenoviridae (Human adenoviruses A-F and murine adenovirus), Flaviviridae (Bovine viral diarrhea virus, TBE, Yellow fever virus, Dengue viruses 1-4, WNV and hepatitis C virus), Paramyxoviridae (Pneumonia virus of mice, Sendai virus, and Simian parainfluenza virus 5), Togaviridae (Western equine encephalomyelitis virus), Picornaviridae (Polio (types 1-13), Human hepatitis A, Human coxsackievirus, Human cardiovirus, Human rhinovirus and Bovine rhinovirus), Reoviridae (Mouse rotavirus, reovirus type 3 and Colorado tick fever virus), and Rhabdoviridae (vesicular stomatitis virus).

BRIEF DESCRIPTION OF THE DRAWINGS

[42] The foregoing summary of the invention, as well as the following detailed description of the invention, is better understood when read in conjunction with the accompanying drawings which are included by way of example and not by way of limitation.

[43] **Figure 1:** process diagram illustrating a representative primer pair selection process.

[44] **Figure 2** is a process diagram illustrating an embodiment of the calibration method.

DEFINITIONS

[45] As used herein, the term "abundance" refers to an amount. The amount may be described in terms of concentration which are common in molecular biology such as "copy number," "pfu or plate-forming unit" which are well known to those with ordinary skill. Concentration may be relative to a known standard or may be absolute.

[46] As used herein an "adventitious virus" or "adventitious viral agent" refers to a virus contaminant present within a biological product, including, for example, vaccines, cell lines and other cell-derived products. In some cases, the biological product may provide a favorable environment for the survival of the virus. In some embodiments, the biological products are those useful in various experimental conditions for research in biotechnology and clinical diagnosis or treatment in pharmacology.

[47] As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" also comprises "sample template."

[48] As used herein the term "amplification" refers to a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (i.e., replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (i.e., synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out. Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Qß replicase, MDV-1 RNA is the specific template for the replicase (D.L. Kacian et al., Proc. Natl. Acad. Sci. USA 69:3038 [1972]). Other nucleic acid will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin et al., Nature 228:227 [1970]). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (D.Y. Wu and R. B. Wallace, Genomics

4:560 [1989]). Finally, Taq and Pfu polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H.A. Erlich (ed.), PCR Technology, Stockton Press [1989]).

[49] As used herein, the term "amplification reagents" refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification, excluding primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.).

[50] As used herein, the term "anion exchange functional group" refers to a positively charged functional group capable of binding an anion through an electrostatic interaction. The most well known anion exchange functional groups are the amines, including primary, secondary, tertiary and quaternary amines.

[51] The term "bacteria" or "bacterium" refers to any member of the groups of eubacteria and archaebacteria.

[52] As used herein, a "base composition" is the exact number of each nucleobase (for example, A, T, C and G). For example, amplification of nucleic acid of *Neisseria meningitidis* with a primer pair that produces an amplification product from nucleic acid of 23S rRNA that has a molecular mass (sense strand) of 28480.75124, from which a base composition of A25 G27 C22 T18 is assigned from a list of possible base compositions calculated from the molecular mass using standard known molecular masses of each of the four nucleobases.

[53] As used herein, a "base composition probability cloud" is a representation of the diversity in base composition resulting from a variation in sequence that occurs among different isolates of a given species. The "base composition probability cloud" represents the base composition constraints for each species and is typically visualized using a pseudo four-dimensional plot.

[54] In the context of this invention, a "bioagent" is any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus. Examples of bioagents include, but are not limited, to cells, (including but not limited to human clinical samples, bacterial cells and other pathogens), viruses, fungi, protists, parasites, and pathogenicity markers (including but not limited to: pathogenicity islands, antibiotic resistance genes, virulence factors, toxin genes and other bioregulating compounds).

Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores) and may be encapsulated or bioengineered. In the context of this invention, a "pathogen" is a bioagent which causes a disease or disorder.

[55] As used herein, a "bioagent division" is defined as group of bioagents above the species level and includes but is not limited to, orders, families, classes, clades, genera or other such groupings of bioagents above the species level.

[56] As used herein, the term "bioagent identifying amplicon" refers to a polynucleotide that is amplified from a bioagent in an amplification reaction and which 1) provides sufficient variability to distinguish each individual bioagent and 2) whose molecular mass is amenable to molecular mass determination.

[57] As used herein, the term "biological product" refers to any product originating from an organism. Biological products are often products of processes of biotechnology. Examples of biological products include, but are not limited to: cultured cell lines, cellular components, antibodies, proteins and other cell-derived biomolecules, growth media, growth harvest fluids, natural products and bio-pharmaceutical products.

[58] The terms "biowarfare agent" and "bioweapon" are synonymous and refer to a bacterium, virus, fungus or protozoan that could be deployed as a weapon to cause bodily harm to individuals by military or terrorist groups.

[59] In context of this invention, the term "broad range survey primer pair" refers to a primer pair designed to produce bioagent identifying amplicons across different broad groupings of bioagents. For example, the ribosomal RNA-targeted primer pairs are broad range survey primer pairs.

[60] The term "calibration amplicon" refers to a nucleic acid segment representing an amplification product obtained by amplification of a calibration sequence with a pair of primers designed to produce a bioagent identifying amplicon.

[61] The term "calibration sequence" refers to a polynucleotide sequence to which a given pair of primers hybridizes for the purpose of producing an internal (i.e: included in the reaction) calibration standard amplification product for use in determining the quantity of a bioagent in a sample. The calibration sequence may be expressly added to an amplification reaction, or may already be present in the sample prior to analysis.

[62] The term "clade primer pair" refers to a primer pair designed to produce bioagent identifying amplicons for species belonging to a clade group. A clade primer pair may also be considered as a speciating primer pair.

[63] The term "codon" refers to a set of three adjoined nucleotides (triplet) that codes for an amino acid or a termination signal.

[64] In context of this invention, the term "codon base composition analysis," refers to determination of the base composition of an individual codon by obtaining a bioagent identifying amplicon that includes the codon. The bioagent identifying amplicon will at least include regions of the target nucleic acid sequence to which the primers hybridize for generation of the bioagent identifying amplicon as well as the codon being analyzed, located between the two primer hybridization regions.

[65] As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (i.e., a sequence of nucleotides such as an oligonucleotide or a target nucleic acid) related by the base-pairing rules. For example, for the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T-C-A-5'." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids. Either term may also be used in reference to individual nucleotides, especially within the context of polynucleotides. For example, a particular nucleotide within an oligonucleotide may be noted for its complementarity, or lack thereof, to a nucleotide within another nucleic acid strand, in contrast or comparison to the complementarity between the rest of the oligonucleotide and the nucleic acid strand.

[66] The term "complement of a nucleic acid sequence" as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in "antiparallel association." Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine and 7-deazaguanine. Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs. Where a first oligonucleotide is complementary to a region of a

target nucleic acid and a second oligonucleotide has complementary to the same region (or a portion of this region) a "region of overlap" exists along the target nucleic acid. The degree of overlap will vary depending upon the extent of the complementarity

[67] In context of this invention, the term "division-wide primer pair" refers to a primer pair designed to produce bioagent identifying amplicons within sections of a broad spectrum of bioagents For example, primer pair number 367, a division-wide primer pair, is designed to produce bioagent identifying amplicons for the beta-proteobacteria division of bacteria.

[68] As used herein, the term "concurrently amplifying" used with respect to more than one amplification reaction refers to the act of simultaneously amplifying more than one nucleic acid in a single reaction mixture.

[69] As used herein, the term "drill down primer pair" refers to a primer pair designed to produce bioagent identifying amplicons for identification of sub-species characteristics.

[70] The term "duplex" refers to the state of nucleic acids in which the base portions of the nucleotides on one strand are bound through hydrogen bonding the their complementary bases arrayed on a second strand. The condition of being in a duplex form reflects on the state of the bases of a nucleic acid. By virtue of base pairing, the strands of nucleic acid also generally assume the tertiary structure of a double helix, having a major and a minor groove. The assumption of the helical form is implicit in the act of becoming duplexed.

[71] As used herein, the term "etiology" refers to the causes or origins, of diseases or abnormal physiological conditions.

[72] The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for the production of an RNA having a non-coding function (e.g., a ribosomal or transfer RNA), a polypeptide or a precursor. The RNA or polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or function is retained.

[73] The terms "homology," "homologous" and "sequence identity" refer to a degree of identity. There may be partial homology or complete homology. A partially homologous sequence is one that is less than 100% identical to another sequence. Determination of sequence identity is described in the following example: a primer 20 nucleobases in length which is otherwise identical to another 20 nucleobase primer but having two non-identical residues has 18 of 20 identical residues (18/20 = 0.9 or

90% sequence identity). In another example, a primer 15 nucleobases in length having all residues identical to a 15 nucleobase segment of a primer 20 nucleobases in length would have 15/20 = 0.75 or 75% sequence identity with the 20 nucleobase primer. In context of the present invention, sequence identity is meant to be properly determined when the query sequence and the subject sequence are both described and aligned in the 5' to 3' direction. Sequence alignment algorithms such as BLAST, will return results in two different alignment orientations. In the Plus/Plus orientation, both the query sequence and the subject sequence are aligned in the 5' to 3' direction. On the other hand, in the Plus/Minus orientation, the query sequence is in the 5' to 3' direction while the subject sequence is in the 3' to 5' direction. It should be understood that with respect to the primers of the present invention, sequence identity is properly determined when the alignment is designated Plus/Plus. Sequence identity may also encompass alternate or modified nucleobases that perform in a functionally similar manner to the regular nucleobases adenine, thymine, guanine and cytosine with respect to hybridization and primer extension in amplification reactions. In a non-limiting example, if the 5-propynyl pyrimidines propyne C and/or propyne T replace one or more C or T residues in one primer which is otherwise identical to another primer in sequence and length, the two primers will have 100% sequence identity with each other. In another non-limiting example, Inosine (I) may be used as a replacement for G or T and effectively hybridize to C, A or U (uracil). Thus, if inosine replaces one or more C, A or U residues in one primer which is otherwise identical to another primer in sequence and length, the two primers will have 100% sequence identity with each other. Other such modified or universal bases may exist which would perform in a functionally similar manner for hybridization and amplification reactions and will be understood to fall within this definition of sequence identity.

[74] As used herein, "housekeeping gene" refers to a gene encoding a protein or RNA involved in basic functions required for survival and reproduction of a bioagent. Housekeeping genes include, but are not limited to genes encoding RNA or proteins involved in translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like.

[75] As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is influenced by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, and the T_m of the formed hybrid. "Hybridization" methods involve the annealing of one nucleic acid to another, complementary nucleic acid, i.e., a nucleic acid having a complementary nucleotide sequence. The ability of two polymers of nucleic acid containing complementary sequences to find each other and anneal through base pairing interaction is a well-recognized phenomenon. The initial observations of the "hybridization" process by Marmur and Lane,

Proc. Natl. Acad. Sci. USA 46:453 (1960) and Doty et al., Proc. Natl. Acad. Sci. USA 46:461 (1960) have been followed by the refinement of this process into an essential tool of modern biology.

[76] The term "*in silico*" refers to processes taking place via computer calculations. For example, electronic PCR (ePCR) is a process analogous to ordinary PCR except that it is carried out using nucleic acid sequences and primer pair sequences stored on a computer formatted medium.

[77] As used herein, "intelligent primers" are primers that are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and, upon amplification, yield amplification products which ideally provide enough variability to distinguish individual bioagents, and which are amenable to molecular mass analysis. By the term "highly conserved," it is meant that the sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity among all, or at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of species or strains.

[78] The "ligase chain reaction" (LCR; sometimes referred to as "Ligase Amplification Reaction" (LAR) described by Barany, Proc. Natl. Acad. Sci., 88:189 (1991); Barany, PCR Methods and Applic., 1:5 (1991); and Wu and Wallace, Genomics 4:560 (1989) has developed into a well-recognized alternative method for amplifying nucleic acids. In LCR, four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, that hybridize to the opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation, hybridization and ligation amplify a short segment of DNA. LCR has also been used in combination with PCR to achieve enhanced detection of single-base changes. However, because the four oligonucleotides used in this assay can pair to form two short ligatable fragments, there is the potential for the generation of target-independent background signal. The use of LCR for mutant screening is limited to the examination of specific nucleic acid positions.

[79] The term "locked nucleic acid" or "LNA" refers to a nucleic acid analogue containing one or more 2'-O, 4'-C-methylene-β-D-ribofuranosyl nucleotide monomers in an RNA mimicking sugar conformation. LNA oligonucleotides display unprecedented hybridization affinity toward complementary single-stranded RNA and complementary single- or double-stranded DNA. LNA oligonucleotides induce A-type (RNA-like) duplex conformations.

[80] As used herein, the term "mass-modifying tag" refers to any modification to a given nucleotide which results in an increase in mass relative to the analogous non-mass modified nucleotide. Mass-modifying tags can include heavy isotopes of one or more elements included in the nucleotide such as carbon-13 for example. Other possible modifications include addition of substituents such as iodine or bromine at the 5 position of the nucleobase for example.

[81] The term "mass spectrometry" refers to measurement of the mass of atoms or molecules. The molecules are first converted to ions, which are separated using electric or magnetic fields according to the ratio of their mass to electric charge. The measured masses are used to identity the molecules.

[82] The term "microorganism" as used herein means an organism too small to be observed with the unaided eye and includes, but is not limited to bacteria, virus, protozoans, fungi; and ciliates.

[83] The term "multi-drug resistant" or multiple-drug resistant" refers to a microorganism which is resistant to more than one of the antibiotics or antimicrobial agents used in the treatment of said microorganism.

[84] The term "multiplex PCR" refers to a PCR reaction where more than one primer set is included in the reaction pool allowing 2 or more different DNA targets to be amplified by PCR in a single reaction tube.

[85] The term "non-template tag" refers to a stretch of at least three guanine or cytosine nucleobases of a primer used to produce a bioagent identifying amplicon which are not complementary to the template. A non-template tag is incorporated into a primer for the purpose of increasing the primer-duplex stability of later cycles of amplification by incorporation of extra G-C pairs which each have one additional hydrogen bond relative to an A-T pair.

[86] The term "nucleic acid sequence" as used herein refers to the linear composition of the nucleic acid residues A, T, C or G or any modifications thereof, within an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single or double stranded, and represent the sense or antisense strand

[87] As used herein, the term "nucleobase" is synonymous with other terms in use in the art including "nucleotide," "deoxynucleotide," "nucleotide residue," "deoxynucleotide residue," "nucleotide triphosphate (NTP)," or deoxynucleotide triphosphate (dNTP).

[88] The term "nucleotide analog" as used herein refers to modified or non-naturally occurring nucleotides such as 5-propynyl pyrimidines (i.e., 5-propynyl-dTTP and 5-propynyl-dTCP), 7-deaza purines (i.e., 7-deaza-dATP and 7-deaza-dGTP). Nucleotide analogs include base analogs and comprise modified forms of deoxyribonucleotides as well as ribonucleotides.

[89] The term "oligonucleotide" as used herein is defined as a molecule comprising two or more deoxyribonucleotides or ribonucleotides, preferably at least 5 nucleotides, more preferably at least about 13 to 35 nucleotides. The exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, PCR, or a combination thereof. Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage, an end of an oligonucleotide is referred to as the "5'-end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3'-end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. A first region along a nucleic acid strand is said to be upstream of another region if the 3' end of the first region is before the 5' end of the second region when moving along a strand of nucleic acid in a 5' to 3' direction. All oligonucleotide primers disclosed herein are understood to be presented in the 5' to 3' direction when reading left to right. When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the "upstream" oligonucleotide and the latter the "downstream" oligonucleotide. Similarly, when two overlapping oligonucleotides are hybridized to the same linear complementary nucleic acid sequence, with the first oligonucleotide positioned such that its 5' end is upstream of the 5' end of the second oligonucleotide, and the 3' end of the first oligonucleotide is upstream of the 3' end of the second oligonucleotide, the first oligonucleotide may be called the "upstream" oligonucleotide and the second oligonucleotide may be called the "downstream" oligonucleotide.

[90] In the context of this invention, a "pathogen" is a bioagent which causes a disease or disorder.

[91] As used herein, the terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

[92] The term "peptide nucleic acid" ("PNA") as used herein refers to a molecule comprising bases or base analogs such as would be found in natural nucleic acid, but attached to a peptide backbone rather than the sugar-phosphate backbone typical of nucleic acids. The attachment of the bases to the peptide is such as to allow the bases to base pair with complementary bases of nucleic acid in a manner similar to that of an oligonucleotide. These small molecules, also designated anti gene agents, stop transcript elongation by binding to their complementary strand of nucleic acid (Nielsen, et al. Anticancer Drug Des. 8:53 63).

[93] The term "polymerase" refers to an enzyme having the ability to synthesize a complementary strand of nucleic acid from a starting template nucleic acid strand and free dNTPs.

[94] As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, hereby incorporated by reference, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified." With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of 32P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

[95] The term "polymerization means" or "polymerization agent" refers to any agent capable of facilitating the addition of nucleoside triphosphates to an oligonucleotide. Preferred polymerization means comprise DNA and RNA polymerases.

[96] As used herein, the terms "pair of primers," or "primer pair" are synonymous. A primer pair is used for amplification of a nucleic acid sequence. A pair of primers comprises a forward primer and a reverse primer. The forward primer hybridizes to a sense strand of a target gene sequence to be amplified and primes synthesis of an antisense strand (complementary to the sense strand) using the target sequence as a template. A reverse primer hybridizes to the antisense strand of a target gene sequence to be amplified and primes synthesis of a sense strand (complementary to the antisense strand) using the target sequence as a template.

[97] The primers are designed to bind to highly conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. In some embodiments, the highly conserved sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity, or between about 99-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus design of the primers requires selection of a variable region with appropriate variability to resolve the identity of a given bioagent. Bioagent identifying amplicons are ideally specific to the identity of the bioagent.

[98] Properties of the primers may include any number of properties related to structure including, but not limited to: nucleobase length which may be contiguous (linked together) or non-contiguous (for example, two or more contiguous segments which are joined by a linker or loop moiety), modified or universal nucleobases (used for specific purposes such as for example, increasing hybridization affinity, preventing non-templated adenylation and modifying molecular mass) percent complementarity to a given target sequences.

[99] Properties of the primers also include functional features including, but not limited to, orientation of hybridization (forward or reverse) relative to a nucleic acid template. The coding or sense strand is the strand to which the forward priming primer hybridizes (forward priming orientation) while the reverse priming primer hybridizes to the non-coding or antisense strand (reverse priming orientation). The functional properties of a given primer pair also include the generic template nucleic acid to which the primer pair hybridizes. For example, identification of bioagents can be accomplished at different levels using primers suited to resolution of each individual level of identification. Broad range survey

primers are designed with the objective of identifying a bioagent as a member of a particular division (e.g., an order, family, genus or other such grouping of bioagents above the species level of bioagents). In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species or sub-species level. Other primers may have the functionality of producing bioagent identifying amplicons for members of a given taxonomic genus, clade, species, sub-species or genotype (including genetic variants which may include presence of virulence genes or antibiotic resistance genes or mutations). Additional functional properties of primer pairs include the functionality of performing amplification either singly (single primer pair per amplification reaction vessel) or in a multiplex fashion (multiple primer pairs and multiple amplification reactions within a single reaction vessel).

[100] As used herein, the terms "purified" or "substantially purified" refer to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated. An "isolated polynucleotide" or "isolated oligonucleotide" is therefore a substantially purified polynucleotide.

[101] The term "reverse transcriptase" refers to an enzyme having the ability to transcribe DNA from an RNA template. This enzymatic activity is known as reverse transcriptase activity. Reverse transcriptase activity is desirable in order to obtain DNA from RNA viruses which can then be amplified and analyzed by the methods of the present invention

[102] The term "Ribosomal RNA" or "rRNA" refers to the primary ribonucleic acid constituent of ribosomes. Ribosomes are the protein-manufacturing organelles of cells and exist in the cytoplasm. Ribosomal RNAs are transcribed from the DNA genes encoding them.

[103] The term "sample" in the present specification and claims is used in its broadest sense. On the one hand it is meant to include a specimen or culture (e.g., microbiological cultures). On the other hand, it is meant to include both biological and environmental samples. A sample may include a specimen of synthetic origin. Biological samples may be animal, including human, fluid, solid (e.g., stool) or tissue, as well as liquid and solid food and feed products and ingredients such as dairy items, vegetables, meat and meat by-products, and waste. Biological samples may be obtained from all of the various families of domestic animals, as well as feral or wild animals, including, but not limited to, such animals as ungulates, bear, fish, lagamorphs, rodents, etc. Environmental samples include environmental material such as surface matter, soil, water and industrial samples, as well as samples obtained from food and dairy processing instruments, apparatus, equipment, utensils, disposable and non-disposable items. These examples are not to be construed as limiting the sample types applicable to the present invention. The

term "source of target nucleic acid" refers to any sample that contains nucleic acids (RNA or DNA). Particularly preferred sources of target nucleic acids are biological samples including, but not limited to blood, saliva, cerebral spinal fluid, pleural fluid, milk, lymph, sputum and semen.

[104] As used herein, the term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is often a contaminant. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

[105] A "segment" is defined herein as a region of nucleic acid within a target sequence.

[106] The "self-sustained sequence replication reaction" (3SR) (Guatelli et al., Proc. Natl. Acad. Sci., 87:1874-1878 [1990], with an erratum at Proc. Natl. Acad. Sci., 87:7797 [1990]) is a transcription-based in vitro amplification system (Kwok et al., Proc. Natl. Acad. Sci., 86:1173-1177 [1989]) that can exponentially amplify RNA sequences at a uniform temperature. The amplified RNA can then be utilized for mutation detection (Fahy et al., PCR Meth. Appl., 1:25-33 [1991]). In this method, an oligonucleotide primer is used to add a phage RNA polymerase promoter to the 5′ end of the sequence of interest. In a cocktail of enzymes and substrates that includes a second primer, reverse transcriptase, RNase H, RNA polymerase and ribo- and deoxyribonucleoside triphosphates, the target sequence undergoes repeated rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the area of interest. The use of 3SR to detect mutations is kinetically limited to screening small segments of DNA (e.g., 200-300 base pairs).

[107] As used herein, the term ""sequence alignment"" refers to a listing of multiple DNA or amino acid sequences and aligns them to highlight their similarities. The listings can be made using bioinformatics computer programs.

[108] In context of this invention, the term "speciating primer pair" refers to a primer pair designed to produce a bioagent identifying amplicon with the diagnostic capability of identifying species members of a group of genera or a particular genus of bioagents. Primer pair number 2922, for example, is a speciating primer pair used to identify species members of the bacterial genus Acinetobacter. Primer pair number 352 is a speciating primer pair used to identify species members of the bacterial genera Streptococcus, Enterococcus, Staphylococcus and Bacillus.

[109] In context of this invention, the term "species confirmation primer pair" refers to a primer pair designed to produce a bioagent identifying amplicon with the diagnostic capability to unambiguously produce a unique base composition to identify a particular species of bioagent.

[110] As used herein, a "sub-species characteristic" is a genetic characteristic that provides the means to distinguish two members of the same bioagent species. For example, one viral strain could be distinguished from another viral strain of the same species by possessing a genetic change (e.g., for example, a nucleotide deletion, addition or substitution) in one of the viral genes, such as the RNA-dependent RNA polymerase.

[111] As used herein, the term "target," refers to a nucleic acid sequence or structure to be detected or characterized. Thus, the "target" is sought to be sorted out from other nucleic acid sequences and contains a sequence that has at least partial complementarity with an oligonucleotide primer. The target nucleic acid may comprise single- or double-stranded DNA or RNA. A "segment" is defined as a region of nucleic acid within the target sequence.

[112] The term "template" refers to a strand of nucleic acid on which a complementary copy is built from nucleoside triphosphates through the activity of a template-dependent nucleic acid polymerase. Within a duplex the template strand is, by convention, depicted and described as the "bottom" strand. Similarly, the non-template strand is often depicted and described as the "top" strand.

[113] As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. Several equations for calculating the T_m of nucleic acids are well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: T_m=81.5+0.41(% G+C), when a nucleic acid is in aqueous solution at 1 M NaCl (see e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985). Other references (e.g., Allawi, H. T. & SantaLucia, J., Jr. Thermodynamics and NMR of internal G.T mismatches in DNA. Biochemistry 36, 10581-94 (1997) include more sophisticated computations which take structural and environmental, as well as sequence characteristics into account for the calculation of T_m.

[114] The term "triangulation genotyping analysis" refers to a method of genotyping a bioagent by measurement of molecular masses or base compositions of amplification products, corresponding to bioagent identifying amplicons, obtained by amplification of regions of more than one gene. In this sense, the term "triangulation" refers to a method of establishing the accuracy of information by

comparing three or more types of independent points of view bearing on the same findings. Triangulation genotyping analysis carried out with a plurality of triangulation genotyping analysis primers yields a plurality of base compositions that then provide a pattern or "barcode" from which a species type can be assigned. The species type may represent a previously known sub-species or strain, or may be a previously unknown strain having a specific and previously unobserved base composition barcode indicating the existence of a previously unknown genotype.

[115] As used herein, the term "triangulation genotyping analysis primer pair" is a primer pair designed to produce bioagent identifying amplicons for determining species types in a triangulation genotyping analysis.

[116] The employment of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as "triangulation identification." Triangulation identification is pursued by analyzing a plurality of bioagent identifying amplicons selected within multiple core genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen et al., J. Appl. Microbiol., 1999, 87, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

[117] In the context of this invention, the term "unknown bioagent" may mean either: (i) a bioagent whose existence is known (such as the well known bacterial species *Staphylococcus aureus* for example) but which is not known to be in a sample to be analyzed, or (ii) a bioagent whose existence is not known (for example, the SARS coronavirus was unknown prior to April 2003). For example, if the method for identification of coronaviruses disclosed in commonly owned U.S. Patent Serial No. 10/829,826 (incorporated herein by reference in its entirety) was to be employed prior to April 2003 to identify the SARS coronavirus in a clinical sample, both meanings of "unknown" bioagent are applicable since the SARS coronavirus was unknown to science prior to April, 2003 and since it was not known what bioagent (in this case a coronavirus) was present in the sample. On the other hand, if the method of U.S. Patent Serial No. 10/829,826 was to be employed subsequent to April 2003 to identify the SARS coronavirus in a clinical sample, only the first meaning (i) of "unknown" bioagent would apply since the SARS coronavirus became known to science subsequent to April 2003 and since it was not known what bioagent was present in the sample.

[118] The term "variable sequence" as used herein refers to differences in nucleic acid sequence between two nucleic acids. For example, the genes of two different bacterial species may vary in sequence by the presence of single base substitutions and/or deletions or insertions of one or more

nucleotides. These two forms of the structural gene are said to vary in sequence from one another. In the context of the present invention, "viral nucleic acid" includes, but is not limited to, DNA, RNA, or DNA that has been obtained from viral RNA, such as, for example, by performing a reverse transcription reaction. Viral RNA can either be single-stranded (of positive or negative polarity) or double-stranded.

[119] The term "virus" refers to obligate, ultramicroscopic, parasites incapable of autonomous replication (i.e., replication requires the use of the host cell's machinery). Viruses can survive outside of a host cell but cannon replicate.

[120] The term "wild-type" refers to a gene or a gene product that has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified", "mutant" or "polymorphic" refers to a gene or gene product that displays modifications in sequence and or functional properties (i.e., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

[121] As used herein, a "wobble base" is a variation in a codon found at the third nucleotide position of a DNA triplet. Variations in conserved regions of sequence are often found at the third nucleotide position due to redundancy in the amino acid code.

DETAILED DESCRIPTION OF EMBODIMENTS

A. Bioagent Identifying Amplicons

[122] The present invention provides methods for detection and identification of unknown bioagents using bioagent identifying amplicons. Primers are selected to hybridize to conserved sequence regions of nucleic acids derived from a bioagent, and which bracket variable sequence regions to yield a bioagent identifying amplicon, which can be amplified and which is amenable to molecular mass determination. The molecular mass then provides a means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass or corresponding base composition signature of the amplification product is then matched against a database of molecular masses or base composition signatures. A match is obtained when an experimentally-determined molecular mass or base composition of an analyzed amplification product is compared with known molecular masses or base compositions of known bioagent identifying amplicons and the experimentally determined molecular mass or base composition is the same as the molecular mass or base composition of one of the known bioagent identifying amplicons. Alternatively, the experimentally-determined molecular

mass or base composition may be within experimental error of the molecular mass or base composition of a known bioagent identifying amplicon and still be classified as a match. In some cases, the match may also be classified using a probability of match model such as the models described in U.S. Serial No. 11/073,362, which is commonly owned and incorporated herein by reference in entirety. Furthermore, the method can be applied to rapid parallel multiplex analyses, the results of which can be employed in a triangulation identification strategy. The present method provides rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent detection and identification.

[123] Despite enormous biological diversity, all forms of life on earth share sets of essential, common features in their genomes. Since genetic data provide the underlying basis for identification of bioagents by the methods of the present invention, it is necessary to select segments of nucleic acids which ideally provide enough variability to distinguish each individual bioagent and whose molecular mass is amenable to molecular mass determination.

[124] Unlike bacterial genomes, which exhibit conversation of numerous genes (i.e. housekeeping genes) across all organisms, viruses do not share a gene that is essential and conserved among all virus families. Therefore, viral identification is achieved within smaller groups of related viruses, such as members of a particular virus family or genus. For example, RNA-dependent RNA polymerase is present in all single-stranded RNA viruses and can be used for broad priming as well as resolution within the virus family.

[125] In some embodiments of the present invention, at least one viral nucleic acid segment is amplified in the process of identifying the bioagent. Thus, the nucleic acid segments that can be amplified by the primers disclosed herein and that provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as bioagent identifying amplicons.

[126] In some embodiments of the present invention, bioagent identifying amplicons comprise from about 45 to about 200 nucleobases (i.e. from about 45 to about 200 linked nucleosides), although both longer and short regions may be used. One of ordinary skill in the art will appreciate that the invention embodies compounds of 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178,

179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, and 200 nucleobases in length, or any range therewithin.

[127] It is the combination of the portions of the bioagent nucleic acid segment to which the primers hybridize (hybridization sites) and the variable region between the primer hybridization sites that comprises the bioagent identifying amplicon.

[128] In some embodiments, bioagent identifying amplicons amenable to molecular mass determination which are produced by the primers described herein are either of a length, size or mass compatible with the particular mode of molecular mass determination or compatible with a means of providing a predictable fragmentation pattern in order to obtain predictable fragments of a length compatible with the particular mode of molecular mass determination. Such means of providing a predictable fragmentation pattern of an amplification product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, for example. Thus, in some embodiments, bioagent identifying amplicons are larger than 200 nucleobases and are amenable to molecular mass determination following restriction digestion. Methods of using restriction enzymes and cleavage primers are well known to those with ordinary skill in the art.

[129] In some embodiments, amplification products corresponding to bioagent identifying amplicons are obtained using the polymerase chain reaction (PCR) that is a routine method to those with ordinary skill in the molecular biology arts. Other amplification methods may be used such as ligase chain reaction (LCR), low-stringency single primer PCR, and multiple strand displacement amplification (MDA). These methods are also known to those with ordinary skill.

B. Primers and Primer Pairs

[130] In some embodiments the primers are designed to bind to conserved sequence regions of a bioagent identifying amplicon that flank an intervening variable region and yield amplification products which provide variability sufficient to distinguish each individual bioagent, and which are amenable to molecular mass analysis. In some embodiments, the highly conserved sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity, or between about 99-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus, design of the primers involves selection of a variable region with sufficient variability to resolve the identity of a given bioagent. In some embodiments, bioagent identifying amplicons are specific to the identity of the bioagent.

[131] In some embodiments, identification of bioagents is accomplished at different levels using primers suited to resolution of each individual level of identification. Broad range survey primers are designed with the objective of identifying a bioagent as a member of a particular division (e.g., an order, family, genus or other such grouping of bioagents above the species level of bioagents). In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species or sub-species level.

[132] In some embodiments, drill-down primers are designed with the objective of identifying a bioagent at the sub-species level (including strains, subtypes, variants and isolates) based on sub-species characteristics. Drill-down intelligent primers are not always required for identification at the sub-species level because broad range survey intelligent primers may, in some cases provide sufficient identification resolution to accomplishing this identification objective.

[133] A representative process flow diagram used for primer selection and validation process is outlined in Figure 1. For each group of organisms, candidate target sequences are identified (**200**) from which nucleotide alignments are created (**210**) and analyzed (**220**). Primers are then designed by selecting appropriate priming regions (**230**) to facilitate the selection of candidate primer pairs (**240**). The primer pairs are then subjected to *in silico* analysis by electronic PCR (ePCR) (**300**) wherein bioagent identifying amplicons are obtained from sequence databases such as GenBank or other sequence collections (**310**) and checked for specificity *in silico* (**320**). Bioagent identifying amplicons obtained from GenBank sequences (**310**) can also be analyzed by a probability model which predicts the capability of a given amplicon to identify unknown bioagents such that the base compositions of amplicons with favorable probability scores are then stored in a base composition database (**325**). Alternatively, base compositions of the bioagent identifying amplicons obtained from the primers and GenBank sequences can be directly entered into the base composition database (**330**). Candidate primer pairs (**240**) are validated by testing their ability to hybridize to target nucleic acid by an *in* vitro amplification by a method such as PCR analysis (**400**) of nucleic acid from a collection of organisms (**410**). Amplification products thus obtained are analyzed by gel electrophoresis or by mass spectrometry to confirm the sensitivity, specificity and reproducibility of the primers used to obtain the amplification products (**420**).

[134] Many of the important pathogens, including the organisms of greatest concern as biowarfare agents, have been completely sequenced. This effort has greatly facilitated the design of primers for the detection of unknown bioagents. The combination of broad-range priming with division-wide and drill-down priming has been used very successfully in several applications of the technology, including environmental surveillance for biowarfare threat agents and clinical sample analysis for medically important pathogens.

[135] Synthesis of primers is well known and routine in the art. The primers may be conveniently and routinely made through the well-known technique of solid phase synthesis. Equipment for such synthesis is sold by several vendors including, for example, Applied Biosystems (Foster City, CA). Any other means for such synthesis known in the art may additionally or alternatively be employed.

[136] In some embodiments primers are employed as compositions for use in methods for identification of viral bioagents as follows: a primer pair composition is contacted with nucleic acid (such as, for example, DNA from a DNA virus, or DNA reverse transcribed from the RNA of an RNA virus) of an unknown viral bioagent. The nucleic acid is then amplified by a nucleic acid amplification technique, such as PCR for example, to obtain an amplification product that represents a bioagent identifying amplicon. The molecular mass of each strand of the double-stranded amplification product is determined by a molecular mass measurement technique such as mass spectrometry for example, wherein the two strands of the double-stranded amplification product are separated during the ionization process. In some embodiments, the mass spectrometry is electrospray Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR-MS) or electrospray time of flight mass spectrometry (ESI-TOF-MS). A list of possible base compositions can be generated for the molecular mass value obtained for each strand and the choice of the correct base composition from the list is facilitated by matching the base composition of one strand with a complementary base composition of the other strand. The molecular mass or base composition thus determined is then compared with a database of molecular masses or base compositions of analogous bioagent identifying amplicons for known viral bioagents. A match between the molecular mass or base composition of the amplification product and the molecular mass or base composition of an analogous bioagent identifying amplicon for a known viral bioagent indicates the identity of the unknown bioagent. In some embodiments, the primer pair used is one of the primer pairs of Table 3. In some embodiments, the method is repeated using a different primer pair to resolve possible ambiguities in the identification process or to improve the confidence level for the identification assignment.

[137] In some embodiments, a bioagent identifying amplicon may be produced using only a single primer (either the forward or reverse primer of any given primer pair), provided an appropriate amplification method is chosen, such as, for example, low stringency single primer PCR (LSSP-PCR). Adaptation of this amplification method in order to produce bioagent identifying amplicons can be accomplished by one with ordinary skill in the art without undue experimentation.

[138] In some embodiments, the oligonucleotide primers are broad range survey primers which hybridize to conserved regions of nucleic acid encoding the PB1 gene or the NUC gene, gene of all (or

between 80% and 100%, between 85% and 100%, between 90% and 100% or between 95% and 100%) known adventitious contaminant viruses and produce bioagent identifying amplicons.

[139] In some cases, the molecular mass or base composition of a viral bioagent identifying amplicon defined by a broad range survey primer pair does not provide enough resolution to unambiguously identify a viral bioagent at or below the species level. These cases benefit from further analysis of one or more viral bioagent identifying amplicons generated from at least one additional broad range survey primer pair or from at least one additional division-wide primer pair. The employment of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as triangulation identification.

[140] In other embodiments, the oligonucleotide primers are division-wide primers which hybridize to nucleic acid encoding genes of species within a genus of viruses. In other embodiments, the oligonucleotide primers are drill-down primers which enable the identification of sub-species characteristics. Drill down primers provide the functionality of producing bioagent identifying amplicons for drill-down analyses such as strain typing when contacted with nucleic acid under amplification conditions. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of viral infections. In some embodiments, sub-species characteristics are identified using only broad range survey primers and division-wide and drill-down primers are not used.

[141] In some embodiments, the primers used for amplification hybridize to and amplify genomic DNA, DNA of bacterial plasmids, DNA of DNA viruses or DNA reverse transcribed from RNA of an RNA virus.

[142] In some embodiments, the primers used for amplification hybridize directly to viral RNA and act as reverse transcription primers for obtaining DNA from direct amplification of viral RNA. Methods of amplifying RNA to produce cDNA using reverse transcriptase are well known to those with ordinary skill in the art and can be routinely established without undue experimentation.

[143] In some embodiments, various computer software programs may be used to aid in design of primers for amplification reactions such as *Primer Premier 5* (Premier Biosoft, Palo Alto, CA) or *OLIGO* Primer Analysis Software (Molecular Biology Insights, Cascade, CO). These programs allow the user to input desired hybridization conditions such as melting temperature of a primer-template duplex for example. In some embodiments, an *in silico* PCR search algorithm, such as (ePCR) is used to analyze primer specificity across a plurality of template sequences which can be readily obtained from public sequence databases such as GenBank for example. An existing RNA structure search algorithm (Macke et

al., Nucl. Acids Res., 2001, 29, 4724-4735, which is incorporated herein by reference in its entirety) has been modified to include PCR parameters such as hybridization conditions, mismatches, and thermodynamic calculations (SantaLucia, Proc. Natl. Acad. Sci. U.S.A., 1998, 95, 1460-1465, which is incorporated herein by reference in its entirety). This also provides information on primer specificity of the selected primer pairs. In some embodiments, the hybridization conditions applied to the algorithm can limit the results of primer specificity obtained from the algorithm. In some embodiments, the melting temperature threshold for the primer template duplex is specified to be 35 °C or a higher temperature. In some embodiments the number of acceptable mismatches is specified to be seven mismatches or less. In some embodiments, the buffer components and concentrations and primer concentrations may be specified and incorporated into the algorithm, for example, an appropriate primer concentration is about 250 nM and appropriate buffer components are 50 mM sodium or potassium and 1.5 mM Mg^{2+}.

[144] One with ordinary skill in the art of design of amplification primers will recognize that a given primer need not hybridize with 100% complementarity in order to effectively prime the synthesis of a complementary nucleic acid strand in an amplification reaction. Moreover, a primer may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event. (e.g., for example, a loop structure or a hairpin structure). The primers of the present invention may comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 99% sequence identity with any of the primers listed in Table 3. Thus, in some embodiments of the present invention, an extent of variation of 70% to 100%, or any range therewithin, of the sequence identity is possible relative to the specific primer sequences disclosed herein. Determination of sequence identity is described in the following example: a primer 20 nucleobases in length which is identical to another 20 nucleobase primer having two non-identical residues has 18 of 20 identical residues (18/20 = 0.9 or 90% sequence identity). In another example, a primer 15 nucleobases in length having all residues identical to a 15 nucleobase segment of primer 20 nucleobases in length would have 15/20 = 0.75 or 75% sequence identity with the 20 nucleobase primer.

[145] Percent homology, sequence identity or complementarity, can be determined by, for example, the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison WI), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489). In some embodiments, complementarity of primers with respect to the conserved priming regions of viral nucleic acid is between about 70% and about 75% 80%. In other embodiments, homology, sequence identity or complementarity, is between about 75% and about 80%. In yet other embodiments, homology, sequence identity or complementarity, is at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or is 100%.

[146] In some embodiments, the primers described herein comprise at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 98%, or at least 99%, or 100% (or any range therewithin) sequence identity with the primer sequences specifically disclosed herein.

[147] One with ordinary skill is able to calculate percent sequence identity or percent sequence homology and able to determine, without undue experimentation, the effects of variation of primer sequence identity on the function of the primer in its role in priming synthesis of a complementary strand of nucleic acid for production of an amplification product of a corresponding bioagent identifying amplicon.

[148] In one embodiment, the primers are at least 13 nucleobases in length. In another embodiment, the primers are less than 36 nucleobases in length.

[149] In some embodiments of the present invention, the oligonucleotide primers are 13 to 35 nucleobases in length (13 to 35 linked nucleotide residues). These embodiments comprise oligonucleotide primers 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 or 35 nucleobases in length, or any range therewithin. The present invention contemplates using both longer and shorter primers. Furthermore, the primers may also be linked to one or more other desired moieties, including, but not limited to, affinity groups, ligands, regions of nucleic acid that are not complementary to the nucleic acid to be amplified, labels, etc. Primers may also form hairpin structures. For example, hairpin primers may be used to amplify short target nucleic acid molecules. The presence of the hairpin may stabilize the amplification complex (see e.g., TAQMAN MicroRNA Assays, Applied Biosystems, Foster City, California).

[150] In some embodiments, any oligonucleotide primer pair may have one or both primers with less then 70% sequence homology with a corresponding member of any of the primer pairs of Table 3 if the primer pair has the capability of producing an amplification product corresponding to a bioagent identifying amplicon. In other embodiments, any oligonucleotide primer pair may have one or both primers with a length greater than 35 nucleobases if the primer pair has the capability of producing an amplification product corresponding to a bioagent identifying amplicon.

[151] In some embodiments, the function of a given primer may be substituted by a combination of two or more primers segments that hybridize adjacent to each other or that are linked by a nucleic acid

loop structure or linker which allows a polymerase to extend the two or more primers in an amplification reaction.

[152] In some embodiments, the primer pairs used for obtaining bioagent identifying amplicons are the primer pairs of Table 3. In other embodiments, other combinations of primer pairs are possible by combining certain members of the forward primers with certain members of the reverse primers. An example can be seen in Table 3 for three primer pair combinations of forward primer POL_NC003461_2253_2279_F (SEQ ID NO: 45), with the reverse primers POL_NC003461_2302_2329_R (SEQ ID NO: 315), POL_NC003461_2320_2349_R, or (SEQ ID NO: 200), POL_NC003461_2320_2352_R (SEQ ID NO: 326). Arriving at a favorable alternate combination of primers in a primer pair depends upon the properties of the primer pair, most notably the size of the bioagent identifying amplicon that would be produced by the primer pair, which should be between about 45 to about 150 nucleobases in length. Alternatively, a bioagent identifying amplicon longer than 150 nucleobases in length could be cleaved into smaller segments by cleavage reagents such as chemical reagents, or restriction enzymes, for example.

[153] In some embodiments, the primers are configured to amplify nucleic acid of a bioagent to produce amplification products that can be measured by mass spectrometry and from whose molecular masses candidate base compositions can be readily calculated.

[154] In some embodiments, any given primer comprises a modification comprising the addition of a non-templated T residue to the 5' end of the primer (i.e., the added T residue does not necessarily hybridize to the nucleic acid being amplified). The addition of a non-templated T residue has an effect of minimizing the addition of non-templated adenosine residues as a result of the non-specific enzyme activity of *Taq* polymerase (Magnuson et al., Biotechniques, 1996, 21, 700-709), an occurrence which may lead to ambiguous results arising from molecular mass analysis.

[155] In some embodiments of the present invention, primers may contain one or more universal bases. Because any variation (due to codon wobble in the 3rd position) in the conserved regions among species is likely to occur in the third position of a DNA (or RNA) triplet, oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind to more than one nucleotide, referred to herein as a "universal nucleobase." For example, under this "wobble" pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal nucleobases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes et al., Nucleosides and Nucleotides, 1995, 14, 1001-1003), the degencrate nucleotides dP or dK (Hill *et al.*), an acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot et al., Nucleosides and

Nucleotides, 1995, 14, 1053-1056) or the purine analog 1-(2-deoxy-β-D-ribofuranosyl)-imidazole-4-carboxamide (Sala et al., Nucl. Acids Res., 1996, 24, 3302-3306).

[156] In some embodiments, to compensate for the somewhat weaker binding by the wobble base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs that bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, 5-propynyluracil which binds to adenine and 5-propynylcytosine and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is commonly owned and incorporated herein by reference in its entirety. Propynylated primers are described in U.S Pre-Grant Publication No. 2003-0170682, which is also commonly owned and incorporated herein by reference in its entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096, each of which is incorporated herein by reference in its entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

[157] In some embodiments, for broad priming of rapidly evolving RNA viruses, primer hybridization is enhanced using primers containing 5-propynyl deoxy-cytidine and deoxy-thymidine nucleotides. These modified primers offer increased affinity and base pairing selectivity.

[158] In some embodiments, non-template primer tags are used to increase the melting temperature (T_m) of a primer-template duplex in order to improve amplification efficiency. A non-template tag is at least three consecutive A or T nucleotide residues on a primer which are not complementary to the template. In any given non-template tag, A can be replaced by C or G and T can also be replaced by C or G. Although Watson-Crick hybridization is not expected to occur for a non-template tag relative to the template, the extra hydrogen bond in a G-C pair relative to an A-T pair confers increased stability of the primer-template duplex and improves amplification efficiency for subsequent cycles of amplification when the primers hybridize to strands synthesized in previous cycles.

[159] In other embodiments, propynylated tags may be used in a manner similar to that of the non-template tag, wherein two or more 5-propynylcytidine or 5-propynyluridine residues replace template matching residues on a primer. In other embodiments, a primer contains a modified internucleoside linkage such as a phosphorothioate linkage, for example.

[160] In some embodiments, the primers contain mass-modifying tags. Reducing the total number of possible base compositions of a nucleic acid of specific molecular weight provides a means of avoiding a

persistent source of ambiguity in determination of base composition of amplification products. Addition of mass-modifying tags to certain nucleobases of a given primer will result in simplification of *de novo* determination of base composition of a given bioagent identifying amplicon from its molecular mass.

[161] In some embodiments of the present invention, the mass modified nucleobase comprises one or more of the following: for example, 7-deaza-2'-deoxyadenosine-5-triphosphate, 5-iodo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxyuridine-5'-triphosphate, 5-bromo-2'-deoxycytidine-5'-triphosphate, 5-iodo-2'-deoxycytidine-5'-triphosphate, 5-hydroxy-2'-deoxyuridine-5'-triphosphate, 4-thiothymidine-5'-triphosphate, 5-aza-2'-deoxyuridine-5'-triphosphate, 5-fluoro-2'-deoxyuridine-5'-triphosphate, O6-methyl-2'-deoxyguanosine-5'-triphosphate, N2-methyl-2'-deoxyguanosine-5'-triphosphate, 8-oxo-2'-deoxyguanosine-5'-triphosphate or thiothymidine-5'-triphosphate. In some embodiments, the mass-modified nucleobase comprises ^{15}N or ^{13}C or both ^{15}N and ^{13}C.

[162] In some embodiments, multiplex amplification is performed where multiple bioagent identifying amplicons are amplified with a plurality of primer pairs. The advantages of multiplexing are that fewer reaction containers (for example, wells of a 96- or 384-well plate) are needed for each molecular mass measurement, providing time, resource and cost savings because additional bioagent identification data can be obtained within a single analysis. Multiplex amplification methods are well known to those with ordinary skill and can be developed without undue experimentation. However, in some embodiments, one useful and non-obvious step in selecting a plurality candidate bioagent identifying amplicons for multiplex amplification is to ensure that each strand of each amplification product will be sufficiently different in molecular mass that mass spectral signals will not overlap and lead to ambiguous analysis results. In some embodiments, a 10 Da difference in mass of two strands of one or more amplification products is sufficient to avoid overlap of mass spectral peaks.

[163] In some embodiments, as an alternative to multiplex amplification, single amplification reactions can be pooled before analysis by mass spectrometry. In these embodiments, as for multiplex amplification embodiments, it is useful to select a plurality of candidate bioagent identifying amplicons to ensure that each strand of each amplification product will be sufficiently different in molecular mass that mass spectral signals will not overlap and lead to ambiguous analysis results.

C Determination of Molecular Mass of Bioagent Identifying Amplicons

[164] In some embodiments, the molecular mass of a given bioagent identifying amplicon is determined by mass spectrometry. Mass spectrometry has several advantages, not the least of which is high bandwidth characterized by the ability to separate (and isolate) many molecular peaks across a broad range of mass to charge ratio (m/z). Thus mass spectrometry is intrinsically a parallel detection scheme

without the need for radioactive or fluorescent labels, since every amplification product is identified by its molecular mass. The current state of the art in mass spectrometry is such that less than femtomole quantities of material can be readily analyzed to afford information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons.

[165] In some embodiments, intact molecular ions are generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but are not limited to, electrospray ionization (ES), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a single mass spectrum affords an estimate of molecular mass of the bioagent identifying amplicon. Electrospray ionization mass spectrometry (ESI-MS) is particularly useful for very high molecular weight polymers such as proteins and nucleic acids having molecular weights greater than 10 kDa, since it yields a distribution of multiply-charged molecules of the sample without causing a significant amount of fragmentation.

[166] The mass detectors used in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), time of flight (TOF), ion trap, quadrupole, magnetic sector, Q-TOF, and triple quadrupole.

D. Base Compositions of Bioagent Identifying Amplicons

[167] Although the molecular mass of amplification products obtained using intelligent primers provides a means for identification of bioagents, conversion of molecular mass data to a base composition signature is useful for certain analyses. As used herein, "base composition" is the exact number of each nucleobase (A, T, C and G) determined from the molecular mass of a bioagent identifying amplicon. In some embodiments, a base composition provides an index of a specific organism. Base compositions can be calculated from known sequences of known bioagent identifying amplicons and can be experimentally determined by measuring the molecular mass of a given bioagent identifying amplicon, followed by determination of all possible base compositions which are consistent with the measured molecular mass within acceptable experimental error. The following example illustrates determination of base composition from an experimentally obtained molecular mass of a 46-mer amplification product originating at position 1337 of the 16S rRNA of *Bacillus anthracis*. The forward and reverse strands of the amplification product have measured molecular masses of 14208 and 14079 Da, respectively. The

possible base compositions derived from the molecular masses of the forward and reverse strands for the *B. anthracis* products are listed in Table 1.

Table 1

Possible Base Compositions for *B. anthracis* 46mer Amplification Product

Calc. Mass Forward Strand	Mass Error Forward Strand	Base Composition of Forward Strand	Calc. Mass Reverse Strand	Mass Error Reverse Strand	Base Composition of Reverse Strand
14208.2935	0.079520	A1 G17 C10 T18	14079.2624	0.080600	A0 G14 C13 T19
14208.3160	0.056980	A1 G20 C15 T10	14079.2849	0.058060	A0 G17 C18 T11
14208.3386	0.034440	A1 G23 C20 T2	14079.3075	0.035520	A0 G20 C23 T3
14208.3074	0.065560	A6 G11 C3 T26	14079.2538	0.089180	A5 G5 C1 T35
14208.3300	0.043020	A6 G14 C8 T18	14079.2764	0.066640	A5 G8 C6 T27
14208.3525	0.020480	A6 G17 C13 T10	14079.2989	0.044100	A5 G11 C11 T19
14208.3751	0.002060	A6 G20 C18 T2	14079.3214	0.021560	A5 G14 C16 T11
14208.3439	0.029060	A11 G8 C1 T26	14079.3440	0.000980	A5 G17 C21 T3
14208.3665	0.006520	A11 G11 C6 T18	14079.3129	0.030140	A10 G5 C4 T27
14208.3890	0.016020	**A11 G14 C11 T10**	14079.3354	0.007600	A10 G8 C9 T19
14208.4116	0.038560	A11 G17 C16 T2	**14079.3579**	0.014940	**A10 G11 C14 T11**
14208.4030	0.029980	A16 G8 C4 T18	14079.3805	0.037480	A10 G14 C19 T3
14208.4255	0.052520	A16 G11 C9 T10	14079.3494	0.006360	A15 G2 C2 T27
14208.4481	0.075060	A16 G14 C14 T2	14079.3719	0.028900	A15 G5 C7 T19
14208.4395	0.066480	A21 G5 C2 T18	14079.3944	0.051440	A15 G8 C12 T11
14208.4620	0.089020	A21 G8 C7 T10	14079.4170	0.073980	A15 G11 C17 T3
-	-	-	14079.4084	0.065400	A20 G2 C5 T19
-	-	-	14079.4309	0.087940	A20 G5 C10 T13

[168] Among the 16 possible base compositions for the forward strand and the 18 possible base compositions for the reverse strand that were calculated, only one pair (shown in **bold**) are complementary base compositions, which indicates the true base composition of the amplification product. It should be recognized that this logic is applicable for determination of base compositions of any bioagent identifying amplicon, regardless of the class of bioagent from which the corresponding amplification product was obtained.

[169] In some embodiments, assignment of previously unobserved base compositions (also known as "true unknown base compositions") to a given phylogeny can be accomplished via the use of pattern classifier model algorithms. Base compositions, like sequences, vary slightly from strain to strain within species, for example. In some embodiments, the pattern classifier model is the mutational probability model. On other embodiments, the pattern classifier is the polytope model. The mutational probability

model and polytope model are both commonly owned and described in U.S. Patent application Serial No. 11/073,362 which is incorporated herein by reference in entirety.

[170] In one embodiment, it is possible to manage this diversity by building "base composition probability clouds" around the composition constraints for each species. This permits identification of organisms in a fashion similar to sequence analysis. A "pseudo four-dimensional plot" can be used to visualize the concept of base composition probability clouds. Optimal primer design requires optimal choice of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual bioagents. Areas where clouds overlap indicate regions that may result in a misclassification, a problem which is overcome by a triangulation identification process using bioagent identifying amplicons not affected by overlap of base composition probability clouds.

[171] In some embodiments, base composition probability clouds provide the means for screening potential primer pairs in order to avoid potential misclassifications of base compositions. In other embodiments, base composition probability clouds provide the means for predicting the identity of a bioagent whose assigned base composition was not previously observed and/or indexed in a bioagent identifying amplicon base composition database due to evolutionary transitions in its nucleic acid sequence. Thus, in contrast to probe-based techniques, mass spectrometry determination of base composition does not require prior knowledge of the composition or sequence in order to make the measurement.

[172] The present invention provides bioagent classifying information similar to DNA sequencing and phylogenetic analysis at a level sufficient to identify a given bioagent. Furthermore, the process of determination of a previously unknown base composition for a given bioagent (for example, in a case where sequence information is unavailable) has downstream utility by providing additional bioagent indexing information with which to populate base composition databases. The process of future bioagent identification is thus greatly improved as more BCS indexes become available in base composition databases.

E. Triangulation Identification

[173] In some cases, a molecular mass of a single bioagent identifying amplicon alone does not provide enough resolution to unambiguously identify a given bioagent. The employment of more than one bioagent identifying amplicon for identification of a bioagent is herein referred to as "triangulation identification." Triangulation identification is pursued by determining the molecular masses of a plurality of bioagent identifying amplicons selected within a plurality of housekeeping genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or

otherwise engineered bioagents. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen et al., J. Appl. Microbiol., 1999, 87, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

[174] In some embodiments, the triangulation identification process can be pursued by characterization of bioagent identifying amplicons in a massively parallel fashion using the polymerase chain reaction (PCR), such as multiplex PCR where multiple primers are employed in the same amplification reaction mixture, or PCR in multi-well plate format wherein a different and unique pair of primers is used in multiple wells containing otherwise identical reaction mixtures. Such multiplex and multi-well PCR methods are well known to those with ordinary skill in the arts of rapid throughput amplification of nucleic acids. In other related embodiments, one PCR reaction per well or container may be carried out, followed by an amplicon pooling step wherein the amplification products of different wells are combined in a single well or container which is then subjected to molecular mass analysis. The combination of pooled amplicons can be chosen such that the expected ranges of molecular masses of individual amplicons are not overlapping and thus will not complicate identification of signals.

F. Codon Base Composition Analysis

[175] In some embodiments of the present invention, one or more nucleotide substitutions within a codon of a gene of an infectious organism confer drug resistance upon an organism which can be determined by codon base composition analysis. The organism can be a bacterium, virus, fungus or protozoan.

[176] In some embodiments, the amplification product containing the codon being analyzed is of a length of about 35 to about 150 nucleobases. The primers employed in obtaining the amplification product can hybridize to upstream and downstream sequences directly adjacent to the codon, or can hybridize to upstream and downstream sequences one or more sequence positions away from the codon. The primers may have between about 70% to 100% sequence complementarity with the sequence of the gene containing the codon being analyzed.

[177] In some embodiments, the codon base composition analysis is undertaken

[178] In some embodiments, the codon analysis is undertaken for the purpose of investigating genetic disease in an individual. In other embodiments, the codon analysis is undertaken for the purpose of investigating a drug resistance mutation or any other deleterious mutation in an infectious organism such as a bacterium, virus, fungus or protozoan. In some embodiments, the virus is an adventitious virus identified in a biological product.

[179] In some embodiments, the molecular mass of an amplification product containing the codon being analyzed is measured by mass spectrometry. The mass spectrometry can be either electrospray (ESI) mass spectrometry or matrix-assisted laser desorption ionization (MALDI) mass spectrometry. Time-of-flight (TOF) is an example of one mode of mass spectrometry compatible with the analyses of the present invention.

[180] The methods of the present invention can also be employed to determine the relative abundance of drug resistant strains of the organism being analyzed. Relative abundances can be calculated from amplitudes of mass spectral signals with relation to internal calibrants. In some embodiments, known quantities of internal amplification calibrants can be included in the amplification reactions and abundances of analyte amplification product estimated in relation to the known quantities of the calibrants.

[181] In some embodiments, upon identification of one or more drug-resistant strains of an infectious organism infecting an individual, one or more alternative treatments can be devised to treat the individual.

G. Determination of the Quantity of a Bioagent

[182] In some embodiments, the identity and quantity of an unknown bioagent can be determined using the process illustrated in Figure 2. Primers **(500)** and a known quantity of a calibration polynucleotide **(505)** are added to a sample containing nucleic acid of an unknown bioagent. The total nucleic acid in the sample is then subjected to an amplification reaction **(510)** to obtain amplification products. The molecular masses of amplification products are determined **(515)** from which are obtained molecular mass and abundance data. The molecular mass of the bioagent identifying amplicon **(520)** provides the means for its identification **(525)** and the molecular mass of the calibration amplicon obtained from the calibration polynucleotide **(530)** provides the means for its identification **(535)**. The abundance data of the bioagent identifying amplicon is recorded **(540)** and the abundance data for the calibration data is recorded **(545)**, both of which are used in a calculation **(550)** which determines the quantity of unknown bioagent in the sample.

[183] A sample comprising an unknown bioagent is contacted with a pair of primers that provide the means for amplification of nucleic acid from the bioagent, and a known quantity of a polynucleotide that comprises a calibration sequence. The nucleic acids of the bioagent and of the calibration sequence are amplified and the rate of amplification is reasonably assumed to be similar for the nucleic acid of the bioagent and of the calibration sequence. The amplification reaction then produces two amplification products: a bioagent identifying amplicon and a calibration amplicon. The bioagent identifying amplicon

and the calibration amplicon should be distinguishable by molecular mass while being amplified at essentially the same rate. Effecting differential molecular masses can be accomplished by choosing as a calibration sequence, a representative bioagent identifying amplicon (from a specific species of bioagent) and performing, for example, a 2-8 nucleobase deletion or insertion within the variable region between the two priming sites. The amplified sample containing the bioagent identifying amplicon and the calibration amplicon is then subjected to molecular mass analysis by mass spectrometry, for example. The resulting molecular mass analysis of the nucleic acid of the bioagent and of the calibration sequence provides molecular mass data and abundance data for the nucleic acid of the bioagent and of the calibration sequence. The molecular mass data obtained for the nucleic acid of the bioagent enables identification of the unknown bioagent and the abundance data enables calculation of the quantity of the bioagent, based on the knowledge of the quantity of calibration polynucleotide contacted with the sample.

[184] In some embodiments, construction of a standard curve where the amount of calibration polynucleotide spiked into the sample is varied provides additional resolution and improved confidence for the determination of the quantity of bioagent in the sample. The use of standard curves for analytical determination of molecular quantities is well known to one with ordinary skill and can be performed without undue experimentation.

[185] In some embodiments, multiplex amplification is performed where multiple bioagent identifying amplicons are amplified with multiple primer pairs which also amplify the corresponding standard calibration sequences. In this or other embodiments, the standard calibration sequences are optionally included within a single vector which functions as the calibration polynucleotide. Multiplex amplification methods are well known to those with ordinary skill and can be performed without undue experimentation.

[186] In some embodiments, the calibrant polynucleotide is used as an internal positive control to confirm that amplification conditions and subsequent analysis steps are successful in producing a measurable amplicon. Even in the absence of copies of the genome of a bioagent, the calibration polynucleotide should give rise to a calibration amplicon. Failure to produce a measurable calibration amplicon indicates a failure of amplification or subsequent analysis step such as amplicon purification or molecular mass determination. Reaching a conclusion that such failures have occurred is in itself, a useful event.

[187] In some embodiments, the calibration sequence is comprised of DNA. In some embodiments, the calibration sequence is comprised of RNA.

[188] In some embodiments, the calibration sequence is inserted into a vector that itself functions as the calibration polynucleotide. In some embodiments, more than one calibration sequence is inserted into the vector that functions as the calibration polynucleotide. Such a calibration polynucleotide is herein termed a "combination calibration polynucleotide." The process of inserting polynucleotides into vectors is routine to those skilled in the art and can be accomplished without undue experimentation. Thus, it should be recognized that the calibration method should not be limited to the embodiments described herein. The calibration method can be applied for determination of the quantity of any bioagent identifying amplicon when an appropriate standard calibrant polynucleotide sequence is designed and used. The process of choosing an appropriate vector for insertion of a calibrant is also a routine operation that can be accomplished by one with ordinary skill without undue experimentation.

H. Identification of Adventitious Viruses

[189] In other embodiments of the present invention, the primer pairs produce bioagent identifying amplicons within stable and highly conserved regions of adventitious contaminant viruses. The advantage to characterization of an amplicon in a highly conserved region is that there is a low probability that the region will evolve past the point of primer recognition, in which case, the amplification step would fail. Such a primer set is thus useful as a broad range survey-type primer. In another embodiment of the present invention, the intelligent primers produce bioagent identifying amplicons in a region which evolves more quickly than the stable region described above. The advantage of characterization bioagent identifying amplicon corresponding to an evolving genomic region is that it is useful for distinguishing emerging strain variants.

[190] The present invention also has significant advantages as a platform for identification of diseases caused by emerging viruses. The present invention eliminates the need for prior knowledge of bioagent sequence to generate hybridization probes. Thus, in another embodiment, the present invention provides a means of determining the etiology of a virus infection when the process of identification of viruses is carried out in a clinical setting and, even when the virus is a new species never observed before. This is possible because the methods are not confounded by naturally occurring evolutionary variations (a major concern for characterization of viruses which evolve rapidly) occurring in the sequence acting as the template for production of the bioagent identifying amplicon. Measurement of molecular mass and determination of base composition is accomplished in an unbiased manner without sequence prejudice.

[191] Another embodiment of the present invention also provides a means of tracking the spread of any species or strain of virus when a plurality of samples obtained from different locations are analyzed by the methods described above in an epidemiological setting. In one embodiment, a plurality of samples from a plurality of different locations is analyzed with primer pairs which produce bioagent identifying

amplicons, a subset of which contains a specific virus. The corresponding locations of the members of the virus-containing subset indicate the spread of the specific virus to the corresponding locations.

[192] Members of the Parvoviridae family are small single stranded DNA viruses with genomes of about 4-5 kilobases long. They can be divided into (i) Dependovirus genus that includes the human helper-dependent adeno-associated virus (AAV) serotypes 1 to 8 and the autonomous avian parvoviruses; the adeno associated viruses (AAV 1-8); (ii) Erythrovirus genus that includes the bovine, chipmunk, and autonomous primate parvoviruses, including human viruses B19 and V9; and (iii) Parvovirus genus that include parvoviruses of other animals and rodents (except for chipmunks), carnivores, and pigs, including murine minute virus (MMV). These parvoviruses can infect several cell types and have been described in clinical samples. AAVs in particular, have been implicated in decreased replication, propagation, and growth of other virus.

[193] Exogenous retroviruses are known to cause various malignant and non-malignant diseases in animals over a wide range of species. These viruses infect most known animals and rodents. Examples include, but are not limited to: Deltaretroidvirus (HTLV 1-4, STLV 1-3), Gammaretrovirus (Murine leukemia virus, PERV), Alpharetrovirus: (Avian leucosis virus and Avian endogenous virus) and Human immunodeficiency viruses 1 and 2).

[194] Polyomaviruses are small double-stranded DNA viruses that can infect several species including humans, primates, rodents, rabbits and birds. Because of their tumorigenic and oncogenic potential, it is important to test for these viruses in cell substrates used for vaccine production.

[195] The Papillomaviridae family of viruses contains more that 150 known species representing varying host-specificity and sequence homology. They have been identified in mammals (humans, simians, bovines, canines, ovines) and in birds. Majority of the human Papillomaviruses (HPVs), including all HPV types traditionally called genital and mucosal HPVs belong to supergroup A. Within supergroup A, there are 11 groups; the most medically important of these are the human Papillomaviruses HPV 16, HPV 18, HPV 31, HPV 45, HPV 11, HPV 6 and HPV 2. Each of these has been reported as "high risk" viruses in the medical literature.

[196] Other viral families which are potential adventitious contaminants include, but are not limited to: Herpesviridae (Human herpesviruses 1 through 8, Bovine herpesvirus, Canine herpesvirus and Simian cytomegalovirus), Hepadnaviridae (Hepatitis B virus), Hepeviridae (Hepatitis E virus), Deltavirus (Hepatitis delta virus), Adenoviridae (Human adenoviruses A-F and murine adenovirus), Flaviviridae (Bovine viral diarrhea virus, TBE, Yellow fever virus, Dengue viruses 1-4, WNV and hepatitis C virus),

Paramyxoviridae (Pneumonia virus of mice, Sendai virus, and Simian parainfluenza virus 5), Togaviridae (Western equine encephalomyelitis virus), Picornaviridae (Polio (types 1-13), Human hepatitis A, Human coxsackievirus, Human cardiovirus, Human rhinovirus and Bovine rhinovirus), Reoviridae (Mouse rotavirus, reovirus type 3 and Colorado tick fever virus), and Rhabdoviridae (vesicular stomatitis virus).

I. Kits

[197] The present invention also provides kits for carrying out the methods described herein. In some embodiments, the kit may comprise a sufficient quantity of one or more primer pairs to perform an amplification reaction on a target polynucleotide from a bioagent to form a bioagent identifying amplicon. In some embodiments, the kit may comprise from one to fifty primer pairs, from one to twenty primer pairs, from one to ten primer pairs, or from two to five primer pairs. In some embodiments, the kit may comprise one or more primer pairs recited in Table 3.

[198] In some embodiments, the kit comprises one or more broad range survey primer(s), division wide primer(s), or drill-down primer(s), or any combination thereof. If a given problem involves identification of a specific bioagent, the solution to the problem may require the selection of a particular combination of primers to provide the solution to the problem. A kit may be designed so as to comprise particular primer pairs for identification of a particular bioagent. A drill-down kit may be used, for example, to distinguish different sub-species types of adventitious contaminant viruses or genetically engineered adventitious contaminant viruses. In some embodiments, the primer pair components of any of these kits may be additionally combined to comprise additional combinations of broad range survey primers and division-wide primers so as to be able to identify the adventitious contaminant virus.

[199] In some embodiments, the kit contains standardized calibration polynucleotides for use as internal amplification calibrants. Internal calibrants are described in commonly owned U.S. Patent Application Serial No: 60/545,425 which is incorporated herein by reference in its entirety.

[200] In some embodiments, the kit comprises a sufficient quantity of reverse transcriptase (if an RNA virus is to be identified for example), a DNA polymerase, suitable nucleoside triphosphates (including alternative dNTPs such as inosine or modified dNTPs such as the 5-propynyl pyrimidines or any dNTP containing molecular mass-modifying tags such as those described above), a DNA ligase, and/or reaction buffer, or any combination thereof, for the amplification processes described above. A kit may further include instructions pertinent for the particular embodiment of the kit, such instructions describing the primer pairs and amplification conditions for operation of the method. A kit may also comprise amplification reaction containers such as microcentrifuge tubes and the like. A kit may also comprise reagents or other materials for isolating bioagent nucleic acid or bioagent identifying amplicons

from amplification, including, for example, detergents, solvents, or ion exchange resins which may be linked to magnetic beads. A kit may also comprise a table of measured or calculated molecular masses and/or base compositions of bioagents using the primer pairs of the kit.

[201] In some embodiments, the kit includes a computer program stored on a computer formatted medium (such as a compact disk or portable USB disk drive, for example) comprising instructions which direct a processor to analyze data obtained from the use of the primer pairs of the present invention. The instructions of the software transform data related to amplification products into a molecular mass or base composition which is a useful concrete and tangible result used in identification and/or classification of bioagents. In some embodiments, the kits of the present invention contain all of the reagents sufficient to carry out one or more of the methods described herein.

[202] While the present invention has been described with specificity in accordance with certain of its embodiments, the following examples serve only to illustrate the invention and are not intended to limit the same. In order that the invention disclosed herein may be more efficiently understood, examples are provided below. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting the invention in any manner.

EXAMPLES

Example 1: Design and Validation of Primers that Define Bioagent Identifying Amplicons for Adventitious contaminant viruses

A. General Process of Primer Design

[203] For design of primers that define adventitious contaminant virus identifying amplicons, a series of adventitious contaminant virus genome segment sequences were obtained, aligned and scanned for regions where pairs of PCR primers would amplify products of about 45 to about 150 nucleotides in length and distinguish species and/or individual strains from each other by their molecular masses or base compositions. A typical process shown in Figure 1 is employed for this type of analysis.

[204] A database of expected base compositions for each primer region was generated using an *in silico* PCR search algorithm, such as (ePCR). An existing RNA structure search algorithm (Macke et al., Nucl. Acids Res., 2001, 29, 4724-4735, which is incorporated herein by reference in its entirety) has been modified to include PCR parameters such as hybridization conditions, mismatches, and thermodynamic calculations (SantaLucia, Proc. Natl. Acad. Sci. U.S.A., 1998, 95, 1460-1465, which is incorporated herein by reference in its entirety). This also provides information on primer specificity of the selected primer pairs.

B. Design of Primers for Identification of Parvoviruses

[205] Primer pairs were designed which broadly target the various genera/species of the Parvovirinae family. Parvoviruses of the most medical concern are: B19, AAV-5 and murine minute virus. Approximately 500 complete Parvovirus genome sequences were obtained from GenBank. These genome sequences (each approximately 5 kilobases long) were aligned and scanned for conserved target regions. Initial survey of the genome alignments revealed very little homology across the three major genera described above. However, regions were identified with significant homologies within each genus that were the target for primer design. In all three genera, the regions of conservation were within two major nodes, one in the rep gene, encoding NS1 protein, and the other in the capsid, cap gene, encoding glycoprotein VP1 protein. This ability to prime across all known instances of species within each of these groups will enable surveillance for known parvoviruses and detection of previously unknown parvoviruses in cell lines.

C. Design of Primers for Identification of Retroviruses

[206] The objective of primer design for this viral family was to obtain primer pairs that would prime and produce retrovirus identifying amplicons for all known members of each of the genus groups and as yet unknown variants. The T-lymphotropic viruses, members of the deltaretrovirus genus, infect primates and cause leukemia and neurologic diseases. These 9 kilobase single stranded RNA viruses are highly transmissible. Primer pairs targeting the transcription activator (tax) gene were designed to broadly prime and resolve all known primate T-lymphotropic viruses including human T-lymphotropic viruses (HTLV-1 and -2 and the newly discovered HTLV-3 and -4), and simian T-lymphotropic viruses (STLV-1, -2 and -3). These primer pairs produce retrovirus identifying amplicons of simian and human T-lymphotropic virus species with distinct base compositions indicating that the primer pairs can yield amplification products which are distinguishable from each other on the basis of molecular masses and base compositions.

D. Design of Primers for Identification of Polyomaviruses

[207] Approximately 200 complete Polyomavirus genome sequences were obtained from GenBank. These genome sequences (approximately 5.3 kilobases long) were aligned to each other using bioinformatics tools built in-house, and scanned for conserved target regions. Initial survey of the genome alignments revealed a high degree of homology between the primate (SV40) and the human viruses (BK and JC), whereas the rest of the species were highly divergent and did not share much sequence homology with these above species. For primer design purposes, SV40, BK and JC viral species were classified as a group. Nine different primer pairs (primer pairs 2549-2557) were designed to this cluster (See Table 3) and are expected to provide redundant detection and resolution of the three important

Polyomavirus species. Most of these primers were targeted to the large T antigen gene of Polyomavirus. Three additional primer pairs (primer pair numbers 2559-2561) were designed to include Lymphotropic papovavirus (LPV, the African green monkey papovavirus). While these new primers were less conserved across any one species, they would nonetheless provide broader coverage of viral detection within this family. Additional primer pairs (RS10-14) targeting the rest of the viral species (murine, avian, bovine, etc.) were also designed. Taken together, these primers would provide complete coverage of all known Polyomaviruses.

[208] All of the polyomavirus primer pairs were tested against multiple target species for performance and sensitivity. To test the performance of these primers, plasmid clones containing full length SV40 (ATCC: VRMC-4) and JC virus (ATCC: VRMC-1) DNA were obtained from ATCC. Plasmid concentrations were determined by optical density measurements and used as an approximate estimate of the amount of input viral DNA template. Serial 10-fold dilutions of the plasmid were used for estimating limits of detection. These were tested against the entire panel of 12 primer pairs (primer pair numbers 2549-2561). The primer pairs were initially tested at 10^{-7} and 10^{-8} fold dilutions of each of the plasmids and showed reliable detections, with the exception of primer 2555. Additional testing of a subset of these primer pairs showed that while several primer pairs were able to detect additional, lower dilutions, some of the primer pairs were unsuccessful at producing polyomavirus identifying amplicons below the 10^{-9} dilution. Based on this initial test, a panel of six primers (primer pair numbers 2550, 2551, 2553, 2554, 2557 and 2559) was chosen for use in cell-line characterization. These primers will be tested against known cell lines containing SV-40 and other Polyomaviruses.

[209] For routine screening of cell lines, it is anticipated that as few as two of the primer pairs described above along with the four primers targeting non-human Polyomavirus can provide complete coverage of all known and potentially novel Polyomaviruses. A nucleic acid segment within the large tumor antigen gene provides opportunities for broad priming across human and simian species due to a codon deletion at position 32 of the simian virus 40, which is exemplified by primer pair number 2555 (SEQ ID NO: 112:207). Murine pneumonotropic virus, African green monkey PyV virus, SV40 virus, BK virus, JC virus, hamster PyV and murine PyV virus can be distinguished from each other on the basis of base compositions of amplification products produced with primer pair number 2560 (SEQ ID NOs: 12:260).

E. Design of Primers for Identification of Papillomaviruses

[210] Broad primer pairs covering a set of important human Papillomaviruses (HPV 16, 18, 31, 45, 11, 6, 2) were designed (primer pair numbers 2533-2536). These belong to different groups, but have all been reported in literature to be "high risk" Covering all of these species broadly combined with group-

specific primer pairs described above would be of great value. Additionally, several primer pairs were designed to cover broadly within a single group or across multiple groups of Papillomaviruses to increase robustness of detection.

[211] All of the primer pairs were tested against a panel of Papillomaviruses obtained from ATCC. The following viruses were obtained as full-length plasmid clones: ATCC **45150D** (HPV-6b); ATCC 45151D (HPV-11); ATCC 45152D (HPV-18); and ATCC 45113D (HPV-16). Two of the broad primer pairs (numbers 2534 and 2536) amplified all four viruses tested at two different dilutions of the plasmids. Primer pair number 2535 (SEQ ID NOs: 28:253) amplified only two of the test isolates, while primer pair 2533 (30:268) did not amplify any of the viruses tested. Based on these initial results, Primer pair numbers 2534 (SEQ ID NOs: 30:267) and 2536 (SEQ ID NOs: 19:267) were selected for further optimization. A series of primer modifications, including, for example, inosine substitutions to overcome potential sequence mismatches were introduced into the forward and reverse primer pairs. Most of the modified primers tested showed improved performance across the test isolates. In addition to the primers broadly targeting the major species, a series of primers targeting Papillomavirus groups, A7, A9 and A10 that account for over 30 different Papillomaviruses were also tested. Table 2 provides the primer pairs used for Papillomavirus identification and indicates isolates tested, target virus groups and major species covered.

Table 2: Primer Pairs Targeting Human Papillomaviruses

Primer Pair Number	Isolates Tested	Target Virus Group	Major Species Covered
2537	HPV-16	Group A9	HPV-16, HPV-31, HPV-33, HPV-35, HPV-52, HPV-58, HPV-67, and RhPV
2539			
2540			
2543	HPV-18	Group A7	HPV-18, HPV-39, HPV-45, HPV-59, HPV-68, and HPV-70
2544			
2545			
2546	HPV-6, HPV-11	Group A10	HPV-6, HPV-11, HPV-13, HPV-44, HPV-55, and PCPV
2547			
2548			
2541	HPV-6, HPV-11,	Groups A1,	>30 different

2542	HPV-18	A7, A8, A10 and A11	Papilloma viruses

F. **Validation of Primer Pairs Designed for Identification of Papillomaviruses**

[212] For additional testing and validation, two different HeLa cell lines infected with HPV-18 were obtained from ATCC (CCL-2 and CCL-2.2). These were tested at limiting dilutions using a subset of the primers tested above. Results are shown below. The primer pairs used for this test included the major human PaV primer pairs, 2534 (SEQ ID NOs: 30:267), 2536 (SEQ ID NOs: 19:267) and 2685 (SEQ ID NOs: 18:272), the multi-group primer 2542 (SEQ ID NOs: 49:218), the Group A7 targeted primers 2544 (SEQ ID NOs: 108:294) and 2545 SEQ ID NOs: 98:193) and the Group A10 primer 2546 (SEQ ID NOs: 64:302).

[213] In addition to testing the performance of the primers on the cell lines, plasmid DNA containing HPV-6b was spiked into the CCL-2 cell line to determine the dynamic range of detection of the two viruses, cell line derived HPV-18 and the plasmid-derived HPV-6b, simultaneously, In all the tests done, the broad primers as well as the Group A7 primers showed detection of HPV-18 in both cell lines at input levels between 1-10 cells per well. At an estimated copy number of approximately 20 HPV-18 genomes per cell, this corresponds to detection sensitivities between 20-200 genomes from cell lines containing papillomavirus sequences. In experiments done with a co-spike of HPV-6b plasmid into these cell lines, the detection ranges were comparable. HPV-6b was spiked in at two different, fixed concentrations of 200 copies and 2000 copies per well and amplified with the broad primer pair number 2534. Simultaneous detection of HPV-6b and HPV-18 was observed when the plasmid DNA was spiked in at 2000 copies into a range of CCL-2 cell concentration from 1000 to 0 per well. HPV-18 was detected in all wells with the exception of the lowest input level (10 cells/well), in the presence of 2000 copies of HPV-6b. HPV-6b (2000 copies) was detected in the presence of HeLa cell loads up to 600 cells/well, with an effective HPV-18 concentration of approximately 12000 genomes/well. In another experiment, a plasmid spike of approximately 200 copies per well was used. In this case, HPV-18 was detected at all test concentrations, including the lowest cell concentration of 10 cells per well. The dynamic range for detection of the two viruses simultaneously is between 5-10 fold at the lower and higher ends, giving an overall dynamic range of ~25 fold for the detection of competing templates in the presence of each other. These experiments indicate that two or more viruses can be simultaneously detected using the same assay.

G. **Primer Pair Compositions for Identification of Adventitious Viruses**

[214] A total of 224 primer pairs were designed. Table 3 which represents a colleetion of primers (sorted by primer pair number) designed to identify adventitious contaminant viruses using the methods described herein. "I" represents inosine. Tp represents propynylated T and Cp represents propynylated

Cp, wherein the propynyl substituent is located at the 5-position of the pyrimidine nucleobase. The primer pair number is an in-house database index number. The forward or reverse primer name shown in Table 3 indicates the gene region of the viral genome to which the primer hybridizes relative to a reference sequence. The forward primer name RVL_X03614_2256_2279_F indicates that the forward primer (_F) hybridizes to residues 2256-2279 of a respirovirus (Paramyxoviridae) sequence (GenBank Accession Number X03614).

Table 3: Primer Pairs for Identification of Adventitious Contaminant Viruses

Primer Pair Number	Forward Primer Name	Forward Sequence	Forward SEQ ID NO:	Reverse Primer Name	Reverse Sequence	Reverse SEQ ID NO:
377	RVL_X03614_2256_2279_F	TAGTGCAATCCATCTAGCAGCTGT	39	RVL_X03614_2302_2324_R	TGATTGTCGCCTTGAACCATTGC	293
378	RVL_X03614_2239_2261_F	TGGACATTCATCTCTATCAGTGC	132	RVL_X03614_2302_2324_R	TGATTGTCGCCTTGAACCATTGC	293
379	PVL_U50363_3176_3199_F	TAGATCCACAAGCTTAGGGTCTG	22	PVL_U50363_3272_3296_R	TGTTGTGCACTTTTGGAGAATATTT	350
380	PVL_U50363_3153_3175_F	TGCTGAATTCGTAACATTGATGA	130	PVL_U50363_3265_3286_R	TTTTGGCGAATATTTTGTTTGG	371
381	MVL_AF266286_1625_1649_F	TAGGGAGACTTTTTGCTAAAATGAC	33	MVL_AF266286_1720_1739_R	TCCTTTGCCATCCCATTGTC	248
382	MVL_AF266286_2033_2054_F	TGTTTGCACGAGGCTAAATGA	170	MVL_AF266286_2122_2139_R	TGGGCAATGAGGGTCACT	318
383	MVL_AF266286_2002_2024_F	TGCCTTAATTGGAGATATGAGAC	126	MVL_AF266286_2035_2054_R	TCATTTAGCCTCTGTGCAAA	234
384	MSVL_AF266286_3529_3548_F	TGTGTCATCTGCGAGTGTGG	168	MSVL_AF266286_3587_3608_R	TTGTCAATATCATCCAGTTGGC	366
385	PNVL_U50363_1285_1301_F	TTTGGACACCCAATGGT	185	PNVL_U50363_1318_1335_R	TACTCTAACAGCATCCAT	199
386	PNVL_U50363_1878_1898_F	TAGAATGTTTGCTATGCAACC	20	PNVL_U50363_1927_1949_R	TTCTCAGCTAACAATTTCTCAGC	357
387	PNVL_U50363_1881_1898_F	TATGTTTGCTATGCAACC	51	PNVL_U50363_1927_1944_R	TGCTAACAATTTCTCAGC	304
388	PNVL_U50363_2669_2687_F	TAGGACCGTGGATAAACAC	27	PNVL_U50363_2743_2763_R	TCTTTCCCCTCTGTATTCTAA	278
389	MNVL_NC004148_24_47_F	TGTTAATGTCTATCTTCCTGACTC	169	MNVL_NC004148_73_96_R	TGAACCAATTGCATTAGTCTCACT	280
390	MNVL_NC004148_30_47_F	TGTCTATCTTCCTGACTC	157	MNVL_NC004148_73_91_R	TAATTGCATTAGTCTCACT	190
391	MNVL_NC004148_1907_1931_F	TAAGTAAGTTCAACCAAGCCTTTAG	5	MNVL_NC004148_1984_2006_R	TGTAACCAGCAGAATAGGCTTTG	332
392	MNVL_NC004148_1917_1931_F	TAACCAAGCCTTTAG	3	MNVL_NC004148_1984_1999_R	TACAGAATAGGCTTTG	191
393	MNVL_NC004148_2389_2414_F	TCAAGAGATCTTCAGTTTATGAGTAA	55	MNVL_NC004148_2468_2488_R	TGTTTATCCATGGTCCCACTC	354
394	MNVL_NC004148_2389_2408_F	TCAAGAGATCTTCAGTTTAT	54	MNVL_NC004148_2479_2498_R	TCTAATATTGTGTTTATCCA	261
395	MNVL_NC004148_2738_2756_F	TGAACATACCAATGCAGTT	104	MNVL_NC004148_2794_2810_R	TCAGGGGTCCTTCTATA	230

409	CHVCP_AY219836_2_22 8_F	TGAGGTAGCCGACGAGGCGGTT	109	CHVCP AY219836_46_70_R	TGAGGATGTGTCCAAGATGCTGCCG	287
410	CHVCP_AY219836_45_6 9_F	TCCCGAGCCATCTTGGCAACATCCTCC	95	CHVCP AY219836_136_160_R	TGCGGGAAAGGCGGGAAGTTGAAGAT	303
411	CHVCP_AY219836_475_ 499_F	TCCCAGGTACTTCCACCCCAAACCTG	73	CHVCP AY219836_577_601_R	TGCCGAGGCCTATGTGGTCGACATT	301
412	CHVCP_AY219836_24_4 8_F	TCAACCACATCAAGGAGTCGGTGTTC	52	CHVCP AY219836_84_108_R	TAGCGAGATGTTGGCGACTCCCCTATT	209
413	CHVRP_AY219836_440_ 464_F	TGGCTGACAACTTTCGAAAGTCAGCGG	139	CHVRP AY219836_495_517_R	TCCGGCCCACTATGACGTGTACA	255
414	CHVRP_AY219836_658_ 683_F	TGGCGATGTATCTACTGAGACTGTGTGA	141	CHVRP AY219836_730_754_R	TGGTAATCAAAATACGTCGGCCAA	325
415	CHVRP_AY219836_775_ 799_F	TGGTACTCCTCCAACTGCTGTCC	146	CHVRP AY219836_843_869_R	TCCGTGGAATGTCGTAGCAGTCGTC	244
990	HIV1_NC001802_7344_ 7364_F	TAGCAGGAAGCACTATGGCG	25	HIV1_NC001802_7413_7439_R	TCAGCAAATTGTCTCTGCTGCACTA	224
991	HIV1_NC001802_7340_ 7361_F	TGAGCAGCAGAAGCACTATGG	111	HIV1_NC001802_7413_7438_R	TAGCAAATTGCTTGCTGTCACACTA	205
992	HIV1_NC001802_4983_ 5014_F	TGTGAATATCGAGCAGGACATAACAA	163	HIV1_NC001802_5104_5127_R	TGGTCTTCCTGGGCCTGTTCCATC	330
993	HIV1_NC001802_1089_ 1117_F	TGTAATCCGCCTATCCCGTAGGAGA AAT	40	HIV1_NC001802_1178_1204_R	TTTGGTCCCTTGTCTTATGTCCAGAATG	370
994	HIV2_NC001722_8414_ 8439_F	TCGAAAAACCTCCAGCCAAGTATCAC	93	HIV2_NC001722_8476_8500_R	TCTAAAGGCACACCCCCATGAATTT	259
995	HIV2_NC001722_8425_ 8452_F	TCAGCCAAGTATCACTCTATCCAGA	65	HIV2_NC001722_8476_8502_R	TGTCTAAAGGCACATCCCCCATGAATTT	341
996	HIV2_NC001722_5050_ 5075_F	TGCCACGGGAGTAGTATGAGAAGCAATGAA	121	HIV2_NC001722_5149_5196_R	TGTCCATTTCCCCCTATTCCTCCCCTCTTT	337
997	HTLV1_NC001436_7221 7245_F	TGCCAGCGGAGTAGTATGAGAAGCAATGAA	121	HTLV2_NC001436_7330_7353_R	TGTCCATTTCCCCCTTCTTTTTAAAATTC ATGC	245
998	HTLV1_NC001436_7094 7118_F	TGCCAATCATCTCATACAACCCCCAA	118	HTLV2_NC001436_7153_7177_R	TGGTCTCGGAAAAGACAGAGGTTCGG	329
999	HTLV1_NC001436_7388 7410_F	TCAGAGCATCGAGATCACTGGGACC	63	HTLV1_NC001436_7489_7516_R	TGAAGCGGAGTCGAGCGATAAGGAAC	289
1000	HTLV1_NC001436_7818 7843_F	GGAGGCCTCCGTTGTCTGCATGTTA	2	HTLV1_NC001436_7925_7947_R	TCGTTTGTAGGGACATCTGCTGAGAAG	257
1001	HTLV1_NC001436_7340 7361_F	TACTCTCACACGGCCTCATAGTTAC	15	HTLV1_NC001436_7404_7428_R	TGGGGCTCATGTTCATTTGTCATC	322
1002	HTLV1_NC001436_7131 7156_F	TCTTTTCCGAGACCCCGACTCC	103	HTLV1_NC001436_7211_7233_R	TGGGAAAGCTGTTAGAGTTACATGC	316
1003	HTLV2_NC001488_8180 8200_F	TCGTTATCCGCTCACGCTCTACAGTTC	99	HTLV2_NC001488_8254_8279_R	TGATGTGATTCCCCGGGTTAAGGAC	291
1004	HTLV2_NC001488_7757 7780_F	TAGAGGCCGATGACAATGGCG	21	HTLV2_NC001488_7840_7861_R	TACTTGGGATTGTTTTGTTGTGGACGG	203
1005	HTLV2_NC001488_7757 7780_F	TCACCAAGGTGCCTCTTAAACGA	58	HTLV2_NC001488_7840_7861_R	TCATTTGTTGTGGTGAGTTCGTTC	233

	_7779_F			R		
1006	HTLV2_NC001488_2435_2456_F	TGGACCATCATTGGAAGGACG	133	HTLV2_NC001488_2516_2540_R	TGGTGTTTGGAGTGGCTATTGGCAG	331
1007	HTLV2_NC001488_3592_3616_F	TTATGTCCTTGGGGTCACCTACTGG	172	HTLV2_NC001488_3680_3704_R	TCTTGCTTTGACATGTTGTGGTGGA	277
1008	HTLV2_NC001488_2880_2904_F	TCCATACTTCGCCTTCACCATTCCC	74	HTLV2_NC001488_2989_3013_R	TGTTGAGGACGGCTGCTAATTGTTG	348
1009	HTLV2_NC001488_1896_1919_F	TTTCGTTGTGGCAAGGTAGGACAC	184	HTLV2_NC001488_1993_2015_R	TTTGAGTTGTTGGGCAGTCCCTTT	369
1010	HTLV2_NC001488_1198_1222_F	TCACCACGCAATGCTTCCCTATCTT	59	HTLV2_NC001488_1268_1291_R	TCCTGCTTGATGGCCTGTAAGTCT	247
1011	HTLV2_NC001488_5735_5758_F	TTGGTCCATGACTCCGACCTTGAA	183	HTLV2_NC001488_5847_5870_R	TGAGGCTGGATCTATCCACGCAAA	288
1012	HTLV2_NC001488_5241_5265_F	TCCGATGCACATTCACGGTTGGTAT	80	HTLV2_NC001488_5332_5356_R	TAGTCGTTGGTCCGTTGTTAGGGAA	215
1013	HCV_NC001433_66_91_F	TCTAGCCATGGCGTTAGTATGAGTGT	101	HCV_NC001433_121_145_R	TGTTCCGCAGACTACTATGGCTCTC	372
1014	HCV_NC001433_66_91_F	TCTAGCCATGGCGTTAGTATGAGTGT	101	HCV_NC001433_146_167_R	TGGCAATTCCGGTGTACTCACC	311
1015	HCV_NC001433_66_91_F	TCTAGCCATGGCGTTAGTATGAGTGT	101	HCV_NC001433_128_153_R	TACTCACCGGTTCCGCAGACCACTAT	197
1016	HCV_NC001433_51_74_F	TTCACGCAGAAAGCGTCTAGCCAT	175	HCV_NC001433_121_145_2_R	TGTTCCGCAGACCACTATGGCTCTC	346
1017	HCV_NC001433_51_73_F	TTCACGCAGAAAGCGTCTAGCCA	174	HCV_NC001433_146_167_R	TGGCAATTCCGGTGTACTCACC	311
1018	HCV_NC001433_51_73_F	TTCACGCAGAAAGCGTCTAGCCA	174	HCV_NC001433_128_153_R	TACTCACCGGTTCCGCAGACCACTAT	197
1019	HCV_NC001433_62_85_F	AGCGTCTAGCCATGGCGTTAGTAT	1	HCV_NC001433_128_153_R	TACTCACCGGTTCCGCAGACCACTAT	197
1020	HCV_NC001433_227_248_F	TCGCAAGACTGCTAGCCGAGTA	94	HCV_NC001433_277_298_R	TCGCAAGCACCCTATCAGGCAG	251
1021	HCV_NC001433_671_698_F	TAGGTCGCGTAATTTGGGTAAGGTCATC	37	HCV_NC001433_720_742_R	TGAATGTACCCCATGAGGTCGGC	284
1022	HCV_NC001433_8598_8623_F	TCCTTCACGGAGGCTATGACTAGGTA	92	HCV_NC001433_8674_8700_R	TCGACACATTGGAGGAGCATGATGTTA	249
1023	WN_NC001563_8365_8390_F	TCAGTGAATATGACTAGCCAGGTGCT	66	WN_NC001563_8434_8463_R	TGTTCCACTTCCCAAGTTTACATCTTCCTC	345
1024	WN_NC001563_8654_8678_F	TGCTCCTCTCAAAACCATGGGACAC	129	WN_NC001563_8749_8771_R	TCAGGAGCTTTCGTGTCCACCTT	227
1025	WN_NC001563_9026_9055_F	TCATCTACAACATGATGGAAAGAGAGAGA	69	WN_NC001563_9101_9121_R	TAGCTCCGAGCCACATGAACC	206
1026	WN_NC001563_10135_10164_F	TGGATAGAGGAGAATGAATGGATGGAAGAC	135	WN_NC001563_10216_10237_R	TCAGGCTGCCACACCAGATGTC	229
1027	WN_NC001563_9016_9043_F	TGTCACACTTGCATTTACAACATGATGG	155	WN_NC001563_9088_9112_R	TCCACATGAACCAAATGGCTCTGCT	235

1028	VH_NC001563_5696_57 21_F	TCAGAGATGGGGGAGTGAAGATTGCCCCTT	56	VH_NC001563_5758_5783_R	TACTCCCGTCTCGTCAGCAACTTCTCTGTT	198
1029	VH_NC001563_5687_57 14_F	TGCCTAGTGTGAAGATGGGGAAATGAG AT	124	PAN_NC001563_5796_5826_R	TGTCTGTGATAACAAAGTCCCAATCATCG TTC	349
1030	HBV_NC001563_128_152_F	TCCCGGCCTGTCAATATGCTAAAACG	83	WH_NC001563_187_212_R	TCGATCAGGCTCCAACATAGCCCCTCT	250
1031	VH_NC001563_5994_60 17_F	TCCCGTCGCAAGTTGGTGATGAGTA	84	VH_NC001563_6079_6104_R	TCGCTATTTGATGTTCTTCAGCATGAT	300
1032	VH_NC001563_3527_35 52_R	TGATTGACCCTTTTTGAGTTGGCCCTT	115	VH_NC001563_3591_3613_R	TCGCTGATCTTGCTCTCCACCTTC	307
1245	HBV_X51970_320_342_F	TCAACCTCCAATCACTCACCAAC	53	HBV_X51970_379_402_R	TATATAGATAAAAGCCCCCAGCAC	216
1246	HBV_X51970_317_341_F	TCCCCAACTCCCAATCACTCACCAA	76	HBV_X51970_379_406_R	TAGGAATATTCATAAAAGCCCCCAGCAC	208
1247	HBV_X51970_311_335_F	TCCCCAGTCCCCAACTCCCAATCACT	96	HBV_X51970_382_410_R	TGAGAGGAATATTCATAAAAGCCCCCAG A	281
1248	HBV_X51970_1376_139 8_F	TCGCTCGCTAGGCTGTGCTGCCAACT	140	HBV_X51970_1412_1436_R	TACGGGACGTTAAACAAAGGACGTCC	195
1249	HBV_X51970_1777_179 8_F	TACTCGGAGGCTGTAGGCATAAATTG GT	13	HBV_X51970_1868_1886_R	TCAGGCACAGCTTGGAGGC	228
1250	HBV_X51970_183_203_F	TACCCCTGCTCGTGTTACAGG	11	HBV_X51970_228_254_R	TGTCTCAGACTCGTGGTGATTGTGGAGA	342
1251	HBV_X51970_180_202_F	TAGGACCCCTGCTCGTGTTACAG	26	HBV_X51970_262_289_R	TGTCTCCCCTAGAAAATTGAGGAGAATC	306
1252	HBV_X51970_374_392_F	TCGCATGTGAGCCTCCTGGCCTT	136	HBV_X51970_433_450_R	TCCCAGAGAACCAACTAGAAGGCTGTAGCC	236
1253	HBV_X51970_368_388_F	TATCGCTGGATGTGTCTGCGG	46	HBV_X51970_424_453_R	TAATCCATGAAGAACCTAACAGAAGGCTGTA GG	189
1254	HBV_X51970_312_335_F	TGTCACGTCCCCAACTCCCAATCACT	117	HBV_X51970_376_398_R	TGATAAAAGCCCCCAGCATCATC	292
2293	HTLV_TAX_GENE_NC_00 1436_7107_7130_F	TCACCTGGGACCCCCATCGATGGAC	60	HTLV_TAX_GENE_NC_001436_7 169_7192_R	TCTCCTGGGTGGGGAAGGAGGGAG	264
2294	HTLV_TAX_GENE_NC_00 1436_7094_7118_F	TCAGAGGATCAGATCACCTGGGACC	63	HTLV_TAX_GENE_NC_001436_7 203_7226_R	TTGGCGGGGTGAGGACCTTGAGGG	355
2295	HTLV_TAX_GENE_NC_00 1436_6997_7019_F	TACCGAGTCTACGTGTTTGGCGACTG TGT	9	HTLV_TAX_GENE_NC_001436_7 109_7139_R	TCCATCGATGGGGTCCCAGGT	240
2408	ARENAS_NC004296_474 _494_F	TGGTGTATGGTGAGATTTTGGGA	149	ARENAS_NC004296_520_540_R	TGGCATGGTGCCCAACTGATT	313
2409	ARENAS_NC004296_474 _494_2_F	TGGGTGTGGTGAGATTTTGGGA	145	ARENAS_NC004296_520_540_2_R	TGGCAATGTGCCCAACTGATT	314
2410	ARENAS_NC004296_931 _953_F	TGTCTTTCAGGAGATGGATGGCC	160	ARENAS_NC004296_982_1002_R	TCGTGTTTCCCCAAGCTCTTCC	344
2411	ARENAS_NC004296_931 _953_2_F	TGTCTTTTCAGGTGATGGATGGCC	161	ARENAS_NC004296_982_1000_R	TGTTTTCCCAAGCCCCC	355
2412	ARENAS_NC004293_459	TGGTGTTGTGAGGGTCTGGGA	150	ARENAS_NC004293_505_527_R	TGCTGGGATGAACTAAACTGATT	308

	_479_F					
2413	ARENAS_NC004293_459_479_F	TGGTGTTGTGAGGGTCTGGGA	150	ARENAS_NC004293_511_534_R	TGGTCAAAGCTGGCATGAACCAAA	327
2414	ARENAS_NC004293_459_479_F	TGGTGTTGTGAGGGTCTGGGA	150	ARENAS_NC004293_511_534_2_R	TGGTCAIAGCIGGCATGAACCAAA	328
2415	ARENAS_NC004293_459_479_F	TGGTGTTGTGAGGGTCTGGGA	150	ARENAS_NC004293_511_534P_R	TGGTCAIAGCIGGCATGAACpCpAAA	328
2416	ARENAS_NC004293_915_937_F	TGTCTCTCTGGAGATGGGTGGCC	159	ARENAS_NC004293_975_994_R	TCAACACTGGTGTTGTCCCA	221
2417	ARENAS_NC004293_915_937_F	TGTCTCTCTGGAGATGGGTGGCC	159	ARENAS_NC004293_975_994P_R	TCAACACTGGTGTpTpGTCpCpCA	221
2418	ARENAS_NC004293_915_937_2_F	TGTCTCTCIGGIGATGGITGGCC	158	ARENAS_NC004293_975_994_R	TCAACACTGGTGTTGTCCCA	221
2419	ARENAS_NC004293_915_937_2_F	TGTCTCTCIGGIGATGGITGGCC	158	ARENAS_NC004293_975_994P_R	TCAACACTGGTGTpTpGTCpCpCA	221
2420	ARENAS_NC004293_915_937P_F	TGTCTCTCIGGIGATGGITpGGCpCp	158	ARENAS_NC004293_975_994P_R	TCAACACTGGTGTpTpGTCpCpCA	221
2423	POL_NC003461_1620_1649_F	TAGTCTCAAAGAGAAAGAGATCAAACAAGA	38	POL_NC003461_1747_1772_R	TCAATCTCTCCCTTAACCATCCCATT	222
2424	POL_NC003461_2107_2138_F	TGTACAATCTATGTAGGAGATCCTTACTGTCC	152	POL_NC003461_2215_2244_R	TGTCCATAATTTCTGGCAATAACCTTCTAT	340
2425	POL_NC003461_2253_2279_F	TATCAGTGCAATCCATCTAGCAGCTGT	45	POL_NC003461_2302_2329_R	TGGCTTGATTGTCTCCTTGAACCATTGC	315
2426	POL_NC003461_2253_2279_F	TATCAGTGCAATCCATCTAGCAGCTGT	45	POL_NC003461_2320_2349_R	TACTCTTGTTGTCACAGCTATAGCTTGATT	200
2427	POL_NC003461_2253_2279_F	TATCAGTGCAATCCATCTAGCAGCTGT	45	POL_NC003461_2320_2352_R	TGGTACTCTTGATGTCACAGCTATAGCTTGATT	326
2428	POL_NC003461_2256_2279_F	TAGTGCAATCCATCTAGCAGCTGT	39	POL_NC003461_2302_2324_R	TGATTGTCGCCTTGAACCATTGC	293
2429	POL_NC003461_2463_2488_F	TGAGACCATCATAAGTAGCAAGATGT	110	POL_NC003461_2497_2523_R	TCTCCCATCATAGTATATCCTTTTGCT	263
2430	POL_NC003461_2935_2960_F	TATAGGGTCATGAATCAAGAACCCGG	41	POL_NC003461_2980_3004_R	TGCATGAATAAGGGTCTGAAGCCCA	299
2431	POL_NC003461_2935_2960_2_F	TATAGGGTCATGAATCAAGAACCIGG	42	POL_NC003461_2980_3004_R	TGCATGAATAAGGGTCTGAAGCCCA	299
2432	POL_NC003461_3644_3668_F	TTGGATCAGCCACTGATGAAAGATC	182	POL_NC003461_3760_3783_R	TGCTTCCATCCACGATATCTCATC	309
2433	POL_NC003461_3759_3782_F	TGATGAGATATCGTGGATGGAAGC	114	POL_NC003461_3880_3912_R	TGCACTAGAAAACTTCATCTGGGTTGCAGTATC	295
2434	POL_NC003461_3759_3782_2_F	TGATGAAATATCGTGGATGGAAGC	113	POL_NC003461_3880_3912_2_R	TGCACTAGAGAATTTCATCTGGGTTGCAGTATC	296
2435	RUBPOL_NC003443_1636_1661_F	TGTGCATCTTACTCACTAAAGGAGAA	164	RUBPOL_NC003443_1708_1734_R	TTCTGCAATTACTTGACACGACCTCAT	358
2436	RUBPOL_NC003443_2067_2093_F	TCAGTTGATGGGCCTACCTCATTTCTT	68	RUBPOL_NC003443_2143_2167_R	TAGGGTCACTTGGAGGATTGAAAGG	210

2437	RUBPOL_NC003443_2133_2156_F	TGTAGGTGATCCCTTCAATCCTCC	154	RUBPOL_NC003443_2251_2275_R	TTGCAGAGATGGAGATCATTGTCCA	359
2438	RUBPOL_NC003443_2237_2261P_F	TATGTCAAAAGCTGTpGGACpAATpGAT	50	RUBPOL_NC003443_2318_2341P_R	TTGCTTGGTTATpCpACpCpCTGAACpCA	360
2439	RUBPOL_NC003443_2249_2276_F	TGTGGACAATGATCTCCATTGCTGCAAT	167	RUBPOL_NC003443_2301_2324_R	TGGACCATACAGGCAACCCGACAA	310
2440	RUBPOL_NC003443_3689_3713_F	TTGCGTCAATGGCTTATATCAAAGG	181	RUBPOL_NC003443_3754_3778_R	TATCCCCGAAAGCCCAGATATATAC	217
2441	PNVL_U50363_2094_2120_F	TACAGATTTCAGCAAGTTCAATCAAGC	7	PNVL_U50363_2161_2181_R	TTGTGCACCATGCAGTTCATC	367
2442	PNVL_U50363_2110_2135_F	TTCAATCAAGCATTTCGGTATGAAAC	173	PNVL_U50363_2161_2181_R	TTGTGCACCATGCAGTTCATC	367
2443	PNVL_U50363_2587_2612_F	TCACGAGATCTGCAGTTTATGAGTAA	62	PNVL_U50363_2677_2704_R	TGAAGTCATCCAGTATAGTGTTTATCCA	283
2444	PNVL_U50363_2581_2612_F	TATATATCACGAGATCTGCAGTTTATGAGTAA	43	PNVL_U50363_2666_2686_R	TGTTTATCCACGGTCCCACTC	353
2445	PNVL_U50363_2583_2609_F	TATATCACGTGATCTGCAGTTTATGAG	44	PNVL_U50363_2677_2704_R	TGAAGTCATCCAGTATAGTGTTTATCCA	283
2446	PNVL_U50363_2660_2687_F	TCCTAAGAGTGGGACCATGGATAAACAC	86	PNVL_U50363_2740_2769_R	TAATAGACTTTCCCCTCTATATTCTAATTC	187
2447	PNVL_U50363_2660_2684_F	TCCTAAGAGTGGGACCATGGATAAA	85	PNVL_U50363_2740_2769_R	TAATAGACTTTCCCCTCTATATTCTAATTC	187
2448	PNVL_U50363_2660_2684_F	TCCTAAGAGTGGGACCATGGATAAA	85	PNVL_U50363_2743_2776_R	TACTGCATAATAGGCTTTCCCCTCTAAATTCTAA	202
2449	PNVL_U50363_2663_2687_F	TAAGAGTGGGACCATGGATAAACAC	4	PNVL_U50363_2740_2769_2_R	TAATAGGCTTTCCCCTCTATATTCTAATTC	188
2450	PNVL_U50363_1868_1894_F	TCAGTGTAGGTAGAATGTTTGCAATGC	67	PNVL_U50363_1918_1946_R	TCAGCTATCAATTTCTCTGCCAATATTTG	225
2451	PNVL_U50363_1874_1898_F	TAGGTAGAATGTTTGCAATGCAACC	36	PNVL_U50363_1918_1946_R	TCAGCTATCAATTTCTCTGCCAATATTTG	225
2452	PNVL_U50363_1872_1894_F	TGTAGGTAGAATGTTTGCAATGC	153	PNVL_U50363_1918_1946_2_R	TCAGCTATCATTTTCTCAGCCAAGATTTG	226
2453	MVL_AF266286_2002_2027_F	TGCCTTAATTGGAGATATGAGACCAT	127	MVL_AF266286_2035_2060_R	TAGATTTCATTTAGCCTCTGTGCAAA	204
2454	MVL_AF266286_2002_2027_2_F	TGCCTGAATTGGAGATATGAGACCAT	125	MVL_AF266286_2125_2144_R	TCGGGAGGGCAATGAGGGTC	254
2467	PNVL_U50363_2660_2684_1F	TCCTAAGAGTGGGACCATGGATAAA	85	PNVL_U50363_2743_2776_1R	TACTGCATAATAGACTTTCCCCTCTATATTCTAA	201
2533	PAV_IMP_NC001526_6222_6253_F	TAGGATGGTGATATGGTTGATACAGGCTTTGG	30	PAV_IMP_NC001526_6321_6355_R	TCTGCAACCATTTGCAAATAATCTGGATATTTGCA	268
2534	PAV_IMP_NC001526_6222_6253_F	TAGGATGGTGATATGGTTGATACAGGCTTTGG	30	PAV_IMP_NC001526_6324_6355_R	TCTGCAACCATTTGCAAATAATCTGGATATTT	267
2535	PAV_IMP_NC001526_6222_6253_2_F	TAGGATGGCGATATGGTTGACACAGGCTTTGG	28	PAV_IMP_NC001526_6324_6355_2_R	TCGGCAGCCATTTGCAAATAATCAGGATATTT	253
2536	PAV_IMP_NC001526_62	TACTGTTATTCAGGATGGTGATATGG	19	PAV_IMP_NC001526_6324_635	TCTGCAACCATTTGCAAATAATCTGGAT	267

	12_6238_F	T		5_R	ATTT	
2537	PAV_A9_NC001526_5632_5655_F	TTCAGATGTCTGTGTGGCGGCCTA	176	PAV_A9_NC001526_5691_5720_R	TACATATTCATCCGTGCTTACAACCTTAGA	192
2538	PAV_A9_NC001526_5632_5655_F	TTCAGATGTCTGTGTGGCGGCCTA	176	PAV_A9_NC001526_5691_5720_R	TACATATTCATCCGTGCTTACAACCTTAGA	192
2539	PAV_A9_NC001526_2688_2715_F	TGGAAATCCTTTTTCTCAAGGACGTGGT	131	PAV_A9_NC001526_2773_2802_R	TAGTATTTTGTCCTGCCACGCATTTAAACG	211
2540	PAV_A9_NC001526_1972_2003_F	TAGATGATAGTGACATTGCATATAAATATGCA	23	PAV_A9_NC001526_2085_2112_R	TTTCTGCTCGTTTATAATGTCTACACAT	368
2541	PAV_A7_NC001357_2011_2036_F	TATGGTGCAGTGGGCATTTGATAATG	49	PAV_A7_NC001357_2096_2121_R	TTGCTTTTTAAAAATGCAGCTGCATT	361
2542	PAV_A7_NC001357_2011_2036_F	TATGGTGCAGTGGGCATTTGATAATG	49	PAV_A7_NC001357_2108_2135_R	TATTTGCCTGCCAATTGCTTTTTAAAAA	218
2543	PAV_A7_NC001357_4504_4530_F	TCCACCTGTGGTTATTGAACCTGT	72	PAV_A7_NC001357_4607_4629_R	TCAAACCCAGAGGTGCCTGTAAA	220
2544	PAV_A7_NC001357_748_767_F	TGACGAACCACACGGTCACA	108	PAV_A7_NC001357_875_895_R	TGCACACAACGGACACACAAA	294
2545	PAV_A7_NC001357_947_966_F	TCGGGATGTAATGGCTGGTT	98	PAV_A7_NC001357_1034_1057_R	TACCATGTCCGAACCTGTATCTGT	193
2546	PAV_A10_NC000904_875_903_F	TCAGGATGGTTTTTGGTAGAGGCTAAGT	64	PAV_A10_NC000904_997_1027_R	TGCCTGTGCTTCCAAGGAATTGTGTGTAATA	302
2547	PAV_A10_NC000904_1000_1027_F	TACACACAATTCCTTGGAAGCACAGGCA	6	PAV_A10_NC000904_1055_1079_R	TTAGGTCCTGCACAGCCGCATAATG	356
2548	PAV_A10_NC000904_1222_1246_F	TGAACTAACGGACAGTGGATATGGC	105	PAV_A10_NC000904_1275_1296_R	TCGCCATGTCTCTCTACCTGCG	252
2549	POLYOMA_NC001669_2685_2713_F	TCAATGTATCTTATCATGTCTGGGTCCCC	57	POLYOMA_NC001669_2774_2796_R	TGAGAGTGGATGGGCAGCCTATG	286
2550	POLYOMA_NC001669_2691_2719_F	TATCTTATCATGTCTGGGTCCCCAGGAAG	47	POLYOMA_NC001669_2774_2796_R	TGAGAGTGGATGGGCAGCCTATG	286
2551	POLYOMA_NC001669_2698_2719_F	TCATGTCTGGGTCCCCTGGAAG	70	POLYOMA_NC001669_2774_2796_R	TGAGAGTGGATGGGCAGCCTATG	286
2552	POLYOMA_NC001669_2726_2749_F	TGTGCCCTCAAAAACCCTGACCTC	166	POLYOMA_NC001669_2774_2796_R	TGAGAGTGGATGGGCAGCCTATG	286
2553	POLYOMA_NC001669_2726_2749_2_F	TGTGCCCTCAAAAACCCTAACCTC	165	POLYOMA_NC001669_2774_2796_R	TGAGAGTGGATGGGCAGCCTATG	286
2554	POLYOMA_NC001669_2767_2788_F	TGCCATTCATAGGCTGCCCATC	122	POLYOMA_NC001669_2899_2917_R	TCTGGAACCCAGCAGTGGA	275
2555	POLYOMA_NC001669_4496_4520_F	TGAGGCCTATAGCAGCTATAGCCTC	112	POLYOMA_NC001669_4583_4606_R	TAGCTGAAATTGCTGCTGGAGAGG	207
2556	POLYOMA_NC001669_4533_4558_F	TCCCTCTACAGTAGCAACGGATGCAA	79	POLYOMA_NC001669_4633_4660_R	TGGGGGACCTAATTGCTACTGTATCTGA	323
2557	POLYOMA_NC001669_4537_4558_F	TCTACAGTAGCAACGGATGCAA	100	POLYOMA_NC001669_4621_4641_R	TGTATCTGAAGCTGCTGCTGC	336
2558	POLYOMA_NC001669_2978_3000_F	TAGGGGCCCAACCCCATTTTCAT	35	POLYOMA_NC001669_3084_3110_R	TCCAGACCCTGCAAAAAATGAGAACAC	238

2559	POLYOMA_NC001669_2967_2986_F	TCACCTTTGCAAAGGGGCCC	61	POLYOMA_NC001669_3087_3115_R	TGGGTCCCTGATCCAACTAGAAATGAAAA	324
2560	POLYOMA_NC001669_3392_3412_F	TACGGTCACAGCTTCCCACAT	12	POLYOMA_NC001669_3424_3446_R	TCTAAATGAGGACCTGACCTGTG	260
2561	POLYOMA_NC001669_3780_3803_F	TCCTAGAAGTAGAGGCAGCATCCA	87	POLYOMA_NC001669_3816_3840_R	TGCTCCAGGAGGTGCAAATCAAAGA	305
2643	HTLV1_NC001436_7387_7410_F	TGGAGGCTCCGTTGTCTGCATGTA	134	HTLV1_NC001436_7489_7516_R	TCGTTTGTAGGGAACATTGGTGAGGAAG	258
2670	PAV_IMP_MOD_NC00152 6_6222_6253_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6321_6355_R	TCTGCAACCATTTGCAAATAATCTGGAT ATTTICA	269
2671	PAV_IMP_MOD_NC00152 6_6222_6253_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6321_6355_2_R	TCTGCAACCATTTGIAAATAATCTGGAT ATTTICA	271
2672	PAV_IMP_MOD_NC00152 6_6222_6253_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6321_6355_3_R	TCTGCAACCATTTIIAAATAATCTGGAT ATTTICA	273
2673	PAV_IMP_MOD_NC00152 6_6222_6253_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6324_6355_R	TCTGCAACCATpTpTpGAAAATAATCTG GATATTT	265
2674	PAV_IMP_MOD_NC00152 6_6222_6253_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6324_6355_R	TCTGCAACCATTTGAAAATAATCTGGAT ATTT	266
2675	PAV_IMP_MOD_NC00152 6_6222_6253_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6324_6355_2_R	TCTGCAACCATTTGIAAATAATCTGGAT ATTT	270
2676	PAV_IMP_MOD_NC00152 6_6222_6253_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6324_6355_3_R	TCTGCAACCATTTIIAIATAATCTGGAT ATTT	272
2677	PAV_IMP_MOD_NC00152 6_6222_6253_2_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	30	PAV_IMP_MOD_NC001526_6324_6355_4_R	TCTGCAACCATTTIIAIATAATCTGGAT ATTT	274
2678	PAV_IMP_MOD_NC00152 6_6222_6253_2_F	TAGGATGGTGATATGGTTGATACAGG CTTTGG	29	PAV_IMP_MOD_NC001526_6324_6355_2_R	TCTGCAACCATTTGIAAATAATCTGGAT ATTT	270
2679	PAV_IMP_MOD_NC00152 6_6222_6253_3_F	TAGGATGGTGATATGGTTGATACAGG ITTTGG	31	PAV_IMP_MOD_NC001526_6324_6355_3_R	TCTGCAACCATTTIIAAATAATCTGGAT ATTT	272
2680	PAV_IMP_MOD_NC00152 6_6222_6253_4_F	TAGGATGGTGATATGGTTGATACTGG ITTTGG	32	PAV_IMP_MOD_NC001526_6324_6355_4_R	TCTGCAACCATTTIIAIATAATCTGGAT ATTT	274
2681	PAV_IMP_MOD_NC00152 6_6212_6238_F	TACTGTTATTCAGGATGGTGATATGG T	19	PAV_IMP_MOD_NC001526_6324_6355_2_R	TCTGCAACCATTTGIAAATAATCTGGAT ATTT	270
2682	PAV_IMP_MOD_NC00152 6_6212_6238_F	TACTGTTATTCAGGATGGTGATATGG T	19	PAV_IMP_MOD_NC001526_6324_6355_3_R	TCTGCAACCATTTIIAAATAATCTGGAT ATTT	272
2683	PAV_IMP_MOD_NC00152 6_6212_6238_F	TACTGTTATTCAGGATGGTGATATGG T	19	PAV_IMP_MOD_NC001526_6324_6355_4_R	TCTGCAACCATTTIIAIATAATCTGGAT ATTT	271
2684	PAV_IMP_MOD_NC00152 6_6212_6238_2_F	TACTGTTATTCAGGATGGTGATATGG T	17	PAV_IMP_MOD_NC001526_6324_6355_2_R	TCTGCAACCATTTGIAAATAATCTGGAT ATTT	270
2685	PAV_IMP_MOD_NC00152 6_6212_6238_3_F	TACTGTTATTCAGGATGGIGATATGG T	18	PAV_IMP_MOD_NC001526_6324_6355_3_R	TCTGCAACCATTTIIAAATAATCTGGAT ATTT	272
2686	PAV_IMP_MOD_NC00152 6_6212_6238_4_F	TACTGTTATTCAGGATGGIGATATGG T	16	PAV_IMP_MOD_NC001526_6324_6355_4_R	TCTGCAACCATTTIIAIATAATCTGGAT ATTT	274
2687	PAV_A9_MOD_NC001526_5632_5655_F	TTCAGATGTCTGTGTGGCIGCCTA	177	PAV_A9_MOD_NC001526_5691_5720_R	TACATATTCATCCGTGCTTACAACCTTA GA	192
2688	PAV_A9_MOD_NC001526	TTCAGATGTCTITGTGGCIGCCTA	178	PAV_A9_MOD_NC001526_5691_	TACATATTCATCCGTGCTTACAACCTTA	192

	_5632_5655_2_F			5720_R	GA	
2689	PAV_A9_MOD_NC001526_2688_2715_F	TGGAAATCCTTTTTCTCAAGGACGTG GT	131	PAV_A9_MOD_NC001526_2773_2802_R	TAGTATTTTGTCCTGCCACICATTTAAA CG	212
2690	PAV_A9_MOD_NC001526_2688_2715_F	TGGAAATCCTTTTTCTCAAGGACGTG GT	131	PAV_A9_MOD_NC001526_2773_2802_2_R	TAGTATTTTGTCCTGCCAIICATTTAAA CG	213
2691	PAV_A9_MOD_NC001526_2688_2715_F	TGGAAATCCTTTTTCTCAAGGACGTG GT	131	PAV_A9_MOD_NC001526_2773_2802_3_R	TAGTATTTTGTCCTGCCIIICATTTAAA CG	214
2692	PAV_A9_MOD_NC001526_1972_2003_F	TAGATGATAGTGAIATIGCATATIAA TATGCA	24	PAV_A9_MOD_NC001526_2085_2112_R	TTTCTGCTCGTTTATAATGTCTACACAT	368
2693	PAV_A7_A10_NC000904_1912_1937_F	TATGGTGCAGTGGGCATTTGATAATG	49	PAV_A7_A10_NC000904_1997_2022_R	TTGCTTTTTAAAAATGCAGIIGCATT	363
2694	PAV_A7_A10_NC000904_1912_1937_F	TATGGTGCAGTGGGCATTTGATAATG	49	PAV_A7_A10_NC000904_1997_2022_2_R	TTGCTTTTTAAAAATGCIIIIGCATT	364
2695	PAV_A7_A10_NC000904_1912_1937_2_F	TATGGTGCAGTGGGCATTTGATAATG	48	PAV_A7_A10_NC000904_1997_2022_R	TTGCTTTTTAAAAATGCAGIIGCATT	362
2696	PAV_A7_A10_NC000904_1912_1937_F	TATGGTGCAGTGGGCATTTGATAATG	49	PAV_A7_A10_NC000904_2009_2036_R	TATTTGCCTGCIIATTGCTITTTAAAAA	219
2807	PYV_NO_MAMMAL_NC001663_3132_3159_F	TAGGGATTTTTGACCCATCTTTTTCT CA	34	PYV_NO_MAMMAL_NC001663_3261_3281_R	TGGGCCTCTCTGCAAAGGAGA	319
2808	PYV_NO_MAMMAL_NC001663_3267_3295_F	TTACAAAGAGGGCCAACGCCATTCTC ATC	171	PYV_NO_MAMMAL_NC001663_3350_3386_R	TGAAAACACAAGATACTTTGGAAACATAC ACAGGAGGT	279
2809	PYV_NO_MAMMAL_NC001663_3578_3598_F	TCCTCCCACAGCAAACATGTG	88	PYV_NO_MAMMAL_NC001663_3689_3716_R	TGTGTCTGTAAAGACTGAAGTTGTTGGA	343
2810	PYV_NO_MAMMAL_NC001663_3561_3586_F	TGTAAATCTAGTGGCTCTCCTCCCAC	151	PYV_NO_MAMMAL_NC001663_3686_3716_R	TGTAACTGTTAAAACTGAGGTTGTTGGA GTG	333
2811	PYV_NO_MAMMAL_NC001663_3267_3294_F	TTGCAAAGAGGCCCAACCCCATTCTC AT	179	PYV_NO_MAMMAL_NC001663_3351_3386_R	TGAGAACACAAGATACTTTGGAAACTIC ACAGGIGG	285
2812	PYV_NO_MAMMAL_NC001663_3267_3294_2_F	TTGCAAAGAGGCCCAACCCCATTITC AT	180	PYV_NO_MAMMAL_NC001663_3351_3386_R	TGAGAACACAAGATACTTTGGAAACTIC ACAGGIGG	285
2813	PYV_NO_MAMMAL_NC001663_3267_3294_F	TTGCAAAGAGGCCCAACCCCATTCTC AT	179	PYV_NO_MAMMAL_NC001663_3372_3404_R	TCCAGACCCAICCAAGAATGAGAACACA AGATA	237
2814	PYV_NO_MAMMAL_NC001663_3267_3294_2_F	TTGCAAAGAGGCCCAACCCCATTITC AT	180	PYV_NO_MAMMAL_NC001663_3372_3404_R	TCCAGACCCAICCAAGAATGAGAACACA AGATA	237
2864	AAV_NS1_NC002077_1005_1028_F	TGCCACGACCGCAAGACCAACAT	119	AAV_NS1_NC002077_1126_1149_R	TGTCATCTTGCCCTCCTCCCACCA	338
2865	AAV_NS1_NC002077_1066_1094_F	TGCGTTAACTGGACCAATGAGAACTT TCC	128	AAV_NS1_NC002077_1126_1149_2_R	TGTCATTTTGCCCTCCTCCCACCA	339
2866	AAV_NS1_NC002077_1020_1046_F	TACCAACATCGCCGAGGCTATIGCCC A	8	AAV_NS1_NC002077_1126_1149_2_R	TGTCATTTTGCCCTCCTCCCACCA	339
2867	AAV_NS1_NC002077_1020_1046_F	TACCAACATCGCCGAGGCTATIGCCC A	8	AAV_NS1_NC002077_1072_1099_R	TGAAGGAAAGTTCTCATTGGTCCAGTT	282
2868	AAV_VP1_NC002077_739_764_F	TGGGTCCTGCCCACCTACAACAACCA	144	AAV_VP1_NC002077_832_859_R	TGTTGAAGTCAAAATACCCCCAGGGGGT	347
2869	AAV_VP1_NC002077_743_764_F	TCCTGCCCACCTACAACAACCA	91	AAV_VP1_NC002077_847_872_R	TGGCAGTGGAATCGGTTGAAGTCAAA	312

2870	AAV_VP1_NC002077_64 8_665_F	TGGTGCCGATGGAGTGGG	148	AAV_VP1_NC002077_680_703_ R	TCCATTTGGAATCGCAATGCCAAT	242
2871	AAV_VP1_NC002077_64 8_665_2_F	TGGTGCCGACGGAGTGGG	147	AAV_VP1_NC002077_680_703_ 2_R	TCCATGGGGAATCGCAATGCCAAT	241
2872	AAV_VP1_NC002077_83 1_851_F	TACCCCCTGGGGGTACTTTGA	10	AAV_VP1_NC002077_886_910_ R	TGTTGTTGATGAGTCTCTGCCAGTC	352
2873	AAV_VP1_NC002077_83 8_867_F	TGGGGGTATTTTGACTTCAACGATT CCAC	142	AAV_VP1_NC002077_886_910_ R	TGTTGTTGATGAGTCTCTGCCAGTC	352
2874	AAV_VP1_NC002077_83 8_867_2_F	TGGGGGTATTTTGACTTCAACIGTT CCAC	143	AAV_VP1_NC002077_886_910_ 2_R	TGTTGTTGATGAGTCITGCCAGTC	351
2875	AAV_VP1_NC002077_67 0_693_F	TCCTCGGGAAATTGGCATTGCGAT	89	AAV_VP1_NC002077_748_772_ R	TGTAGAGGTGGTTGTTGTAGGTGGG	335
2974	EBNA-2_NC007605-36216-37679_780_803_F	TCCTCTCACTCATCAGAGCACCCC	90	EBNA-2_NC007605-36216-37679_853_879_R	TGAGGGGGATAATGGCATAGGAGGAAT	290
2975	EBNA-2_NC007605-36216-37679_2_29_F	TGCCTACATTCTATCTTGCGTTACAT GG	123	EBNA-2_NC007605-36216-37679_70_90_R	TACGGAGAGTGACGGGTTTCC	194
2976	EBNA-2_NC007605-36216-37679_1195_1214_F	TACTCACCAACTCCTGGCCC	14	EBNA-2_NC007605-36216-37679_1270_1292_R	TGGGACTCTGGTTCATGTATTGG	317
2977	EBNA-2_NC007605-36216-37679_1319_1340_F	TCCCCGGTGATTGGTATCCTCC	78	EBNA-2_NC007605-36216-37679_1390_1418_R	TCTTCATCAGAGCTAGGAGATTCTGTTG T	276
2978	EBNA-3A_NC007605-79955-82877_24_41_F	TCCGGCCCTGGATGACAA	82	EBNA-3A_NC007605-79955-82877_94_119_R	TCTATGAGGACATTTTCCCAATCTCC	262
2979	EBNA-3A_NC007605-79955-82877_1318_1337_F	TCGGCGCAAGTCCCAGAACC	97	EBNA-3A_NC007605-79955-82877_1387_1406_R	TACGGGGCCATGCCGTGTTG	196
2980	EBNA-3A_NC007605-79955-82877_1397_1415_F	TGGCCCCGTGTCTGGTAGC	137	EBNA-3A_NC007605-79955-82877_1492_1511_R	TCGGGTACTGGTGCACACGC	256
2981	EBNA-3A_NC007605-79955-82877_1617_1637_F	TGTCCCGACTGTGGCACTTGA	156	EBNA-3A_NC007605-79955-82877_1669_1691_R	TGCATAGCAATCTCAGGAGGTGC	298
2982	EBNA-3B_NC007605-83065-85959_1950_1973_F	TCCCCATCAGACACCTCAGGTGGA	77	EBNA-3B_NC007605-83065-85959_2002_2029_R	TGGGCTGATATGGAATGTGCCCTATCTG	320
2983	EBNA-3B_NC007605-83065-85959_891_911_F	TGAATGGTTCCGCCAGTGCAC	106	EBNA-3B_NC007605-83065-85959_946_968_R	TAAGGAACGGCGACAGGATGCGC	186
2984	EBNA-3B_NC007605-83065-85959_1046_1073_F	TGCCAGATGATCCTATAATTGTTGAG GA	120	EBNA-3B_NC007605-83065-85959_1150_1171_R	TGTACGGCAGTGTTTGCGGTAT	334
2985	EBNA-3B_NC007605-83065-85959_2531_2552_F	TGACACAGGCACCCACGGAATA	107	EBNA-3B_NC007605-83065-85959_2593_2617_R	TCACTCGTTTAGACGCGGAATATC	223

2986	EBNA-3C_NC007605-86083-89135_1284_1307_F	TGTGAAGCGCACAATTGTTAAGAC	162	EBNA-3C_NC007605-86083-89135_1330_1349_R	TCAGGGGTGCTTTGTGCTTC	231
2987	EBNA-3C_NC007605-86083-89135_2549_2573_F	TCCCACATCTGCAATCGGAGACAGG	75	EBNA-3C_NC007605-86083-89135_2620_2642_R	TGGGGAGATGACCATGATGGTGC	321
2988	EBNA-3C_NC007605-86083-89135_1538_1562_F	TGATTGATGTTGAAACCACCGAAGA	116	EBNA-3C_NC007605-86083-89135_1579_1600_R	TCCGATGTGGCTTATTTGGCTG	243
2989	EBNA-1_NC007605-95662-97587_1466_1493_F	TGGCTAGGTGTCACGTAGAAAGGACTAC	138	EBNA-1_NC007605-95662-97587_1510_1533_R	TCCATATACGAACACACCGGCGAC	239
2990	EBNA-1_NC007605-95662-97587_1420_1448_F	TCCAACCAGAAATTTGAGAACATTGCAGA	71	EBNA-1_NC007605-95662-97587_1477_1499_R	TCCTCGGTAGTCCTTTCTACGTG	246
2991	LMP1_S75235-1-1149_1023_1043_F	TCCGCCTTCGATGACAGACGG	81	LMP1_S75235-1-1149_1057_1076_R	TGCAGCGTAGGAAGGTGTGG	297
2992	LMP1_S75235-1-1149_600_620_F	TCTCCCGCACCCTCAACAAGC	102	LMP1_S75235-1-1149_658_679_R	TCAGGTGGTGTCTGCCCTCGTT	232

[215] Table 4 indicates the primer pair name virus identifier for the primer pairs disclosed herein.

Table 4: Primer Pair Name Identifiers for Selected Viruses

Virus Species	Virus Family	Primer Pair Name Virus Identifier	GenBank Accession Numbers
Arenavirus	Arenaviridae	ARENAS	NC_004296
Circovirus	Circoviridae	CRVCP	AY219836
Hepatitis C virus	Flaviviridae	HCV	NC_001433
West Nile virus	Flaviviridae	WN	NC_001563
Hepatitis B virus	Hepadnaviridae	HBV	X51970
Papillomavirus	Papillomaviridae	PAV_IMP	NC_001526
Papillomavirus	Papillomaviridae	PAV_A9	NC_001526
Papillomavirus	Papillomaviridae	PAV_A7	NC_001357
Papillomavirus	Papillomaviridae	PAV_A10	NC_000904
Respirovirus	Paramyxoviridae	RVL	X03614
Pneumovirus	Paramyxoviridae	PVL	U50363
Pneumovirus	Paramyxoviridae	PVNL	U50363
Morbillivirus	Paramyxoviridae	MVL	AF266286
Morbillivirus	Paramyxoviridae	MSVL	AF266286
Metapneumovirus	Paramyxoviridae	MNVL	NC_004148
Respirovirus	Paramyxoviridae	POL	NC_003461
Rubulavirus	Paramyxoviridae	RUBPOL	NC_003443
Human Immunodeficiency virus 1	Retroviridae	HIV1	NC_001802
Human Immunodeficiency virus 2	Retroviridae	HIV-2	NC_001722
Human T-lymphotropic virus 1	Retroviridae	HTLV1	NC_001436
Human T-lymphotropic virus 2	Retroviridae	HTLV2	NC_001488
Human T-cell Lymphotropic Virus	Retroviridae	HTLV_TAX_GENE	NC_001436
Polyomavirus	Polyomaviridae	POLYOMA	NC_001669
Non-mammalian Polyomavirus	Polyomaviridae	PYV_NO_MAMMAL	NC_001663
Dependovirus	Parvoviridae	AAV_NS1	NC_002077
Dependovirus	Parvoviridae	AAV_VP1	NC_002077

Example 2: Sample Preparation and PCR

[216] Samples were processed to obtain viral genomic material using a Qiagen QIAamp Virus BioRobot MDx Kit. Resulting genomic material was amplified using an Eppendorf thermal cycler and

the amplicons were characterized on a Bruker Daltonics MicroTOF instrument. The resulting data was analyzed using GenX software (SAIC, San Diego, CA and Ibis, Carlsbad, CA).

[217] All PCR reactions were assembled in 50 μL reaction volumes in a 96-well microtiter plate format using a Packard MPII liquid handling robotic platform and M.J. Dyad thermocyclers (MJ research, Waltham, MA). The PCR reaction mixture consisted of 4 units of Amplitaq Gold, 1x buffer II (Applied Biosystems, Foster City, CA), 1.5 mM $MgCl_2$, 0.4 M betaine, 800 μM dNTP mixture and 250 nM of each primer. The following typical PCR conditions were used: 95°C for 10 min followed by 8 cycles of 95°C for 30 seconds, 48°C for 30 seconds, and 72°C 30 seconds with the 48°C annealing temperature increasing 0.9°C with each of the eight cycles. The PCR was then continued for 37 additional cycles of 95°C for 15 seconds, 56°C for 20 seconds, and 72°C 20 seconds.

Example 3: Solution Capture Purification of PCR Products for Mass Spectrometry with Ion Exchange Resin-Magnetic Beads

[218] For solution capture of nucleic acids with ion exchange resin linked to magnetic beads, 25 μl of a 2.5 mg/mL suspension of BioClone amine terminated superparamagnetic beads were added to 25 to 50 μl of a PCR (or RT-PCR) reaction containing approximately 10 pM of a typical PCR amplification product. The above suspension was mixed for approximately 5 minutes by vortexing or pipetting, after which the liquid was removed after using a magnetic separator. The beads containing bound PCR amplification product were then washed three times with 50mM ammonium bicarbonate/50% MeOH or 100mM ammonium bicarbonate/50% MeOH, followed by three more washes with 50% MeOH. The bound PCR amplicon was eluted with a solution of 25mM piperidine, 25mM imidazole, 35% MeOH which included peptide calibration standards.

Example 4: Mass Spectrometry and Base Composition Analysis

[219] The ESI-FTICR mass spectrometer is based on a Bruker Daltonics (Billerica, MA) Apex II 70e electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer that employs an actively shielded 7 Tesla superconducting magnet. The active shielding constrains the majority of the fringing magnetic field from the superconducting magnet to a relatively small volume. Thus, components that might be adversely affected by stray magnetic fields, such as CRT monitors, robotic components, and other electronics, can operate in close proximity to the FTICR spectrometer. All aspects of pulse sequence control and data acquisition were performed on a 600 MHz Pentium II data station running Bruker's Xmass software under Windows NT 4.0 operating system. Sample aliquots, typically 15 μl, were extracted directly from 96-well microtiter plates using a CTC HTS PAL autosampler (LEAP Technologies, Carrboro, NC) triggered by the FTICR data station. Samples were injected directly into a

10 μl sample loop integrated with a fluidics handling system that supplies the 100 μl /hr flow rate to the ESI source. Ions were formed via electrospray ionization in a modified Analytica (Branford, CT) source employing an off axis, grounded electrospray probe positioned approximately 1.5 cm from the metalized terminus of a glass desolvation capillary. The atmospheric pressure end of the glass capillary was biased at 6000 V relative to the ESI needle during data acquisition. A counter-current flow of dry N_2 was employed to assist in the desolvation process. Ions were accumulated in an external ion reservoir comprised of an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode, prior to injection into the trapped ion cell where they were mass analyzed. Ionization duty cycles greater than 99% were achieved by simultaneously accumulating ions in the external ion reservoir during ion detection. Each detection event consisted of 1M data points digitized over 2.3 s. To improve the signal-to-noise ratio (S/N), 32 scans were co-added for a total data acquisition time of 74 s.

[220] The ESI-TOF mass spectrometer is based on a Bruker Daltonics MicroTOF™. Ions from the ESI source undergo orthogonal ion extraction and are focused in a reflectron prior to detection. The TOF and FTICR are equipped with the same automated sample handling and fluidics described above. Ions are formed in the standard MicroTOF™ ESI source that is equipped with the same off-axis sprayer and glass capillary as the FTICR ESI source. Consequently, source conditions were the same as those described above. External ion accumulation was also employed to improve ionization duty cycle during data acquisition. Each detection event on the TOF was comprised of 75,000 data points digitized over 75 μs.

[221] The sample delivery scheme allows sample aliquots to be rapidly injected into the electrospray source at high flow rate and subsequently be electrosprayed at a much lower flow rate for improved ESI sensitivity. Prior to injecting a sample, a bolus of buffer was injected at a high flow rate to rinse the transfer line and spray needle to avoid sample contamination/carryover. Following the rinse step, the autosampler injected the next sample and the flow rate was switched to low flow. Following a brief equilibration delay, data acquisition commenced. As spectra were co-added, the autosampler continued rinsing the syringe and picking up buffer to rinse the injector and sample transfer line. In general, two syringe rinses and one injector rinse were required to minimize sample carryover. During a routine screening protocol a new sample mixture was injected every 106 seconds. More recently a fast wash station for the syringe needle has been implemented which, when combined with shorter acquisition times, facilitates the acquisition of mass spectra at a rate of just under one spectrum/minute.

[222] Raw mass spectra were post-calibrated with an internal mass standard and deconvoluted to monoisotopic molecular masses. Unambiguous base compositions were derived from the exact mass measurements of the complementary single-stranded oligonucleotides. Quantitative results are obtained

by comparing the peak heights with an internal PCR calibration standard present in every PCR well at 500 molecules per well. Calibration methods are commonly owned and disclosed in U.S. Provisional Patent Application Serial No. 60/545,425 which is incorporated herein by reference in entirety.

Example 5: *De Novo* Determination of Base Composition of Amplification Products using Molecular Mass Modified Deoxynucleotide Triphosphates

[223] Because the molecular masses of the four natural nucleobases have a relatively narrow molecular mass range (A = 313.058, G = 329.052, C = 289.046, T = 304.046 – See Table 5), a persistent source of ambiguity in assignment of base composition can occur as follows: two nucleic acid strands having different base composition may have a difference of about 1 Da when the base composition difference between the two strands is G ↔ A (-15.994) combined with C ↔ T (+15.000). For example, one 99-mer nucleic acid strand having a base composition of $A_{27}G_{30}C_{21}T_{21}$ has a theoretical molecular mass of 30779.058 while another 99-mer nucleic acid strand having a base composition of $A_{26}G_{31}C_{22}T_{20}$ has a theoretical molecular mass of 30780.052. A 1 Da difference in molecular mass may be within the experimental error of a molecular mass measurement and thus, the relatively narrow molecular mass range of the four natural nucleobases imposes an uncertainty factor.

[224] The present invention provides for a means for removing this theoretical 1 Da uncertainty factor through amplification of a nucleic acid with one mass-tagged nucleobase and three natural nucleobases. The term "nucleobase" as used herein is synonymous with other terms in use in the art including "nucleotide," "deoxynucleotide," "nucleotide residue," "deoxynucleotide residue," "nucleotide triphosphate (NTP)," or deoxynucleotide triphosphate (dNTP).

[225] Addition of significant mass to one of the 4 nucleobases (dNTPs) in an amplification reaction, or in the primers themselves, will result in a significant difference in mass of the resulting amplification product (significantly greater than 1 Da) arising from ambiguities arising from the G ↔ A combined with C ↔ T event (Table 5). Thus, the same the G ↔ A (-15.994) event combined with 5-Iodo-C ↔ T (-110.900) event would result in a molecular mass difference of 126.894. If the molecular mass of the base composition $A_{27}G_{30}$**5-Iodo-**$C_{21}T_{21}$ (33422.958) is compared with $A_{26}G_{31}$**5-Iodo-**$C_{22}T_{20}$, (33549.852) the theoretical molecular mass difference is +126.894. The experimental error of a molecular mass measurement is not significant with regard to this molecular mass difference. Furthermore, the only base composition consistent with a measured molecular mass of the 99-mer nucleic acid is $A_{27}G_{30}$**5-Iodo-**$C_{21}T_{21}$. In contrast, the analogous amplification without the mass tag has 18 possible base compositions.

Table 5: Molecular Masses of Natural Nucleobases and the Mass-Modified Nucleobase 5-Iodo-C and Molecular Mass Differences Resulting from Transitions

Nucleobase	Molecular Mass	Transition	Δ Molecular Mass
A	313.058	A-->T	-9.012
A	313.058	A-->C	-24.012
A	313.058	A-->5-Iodo-C	101.888
A	313.058	A-->G	15.994
T	304.046	T-->A	9.012
T	304.046	T-->C	-15.000
T	304.046	T-->5-Iodo-C	110.900
T	304.046	T-->G	25.006
C	289.046	C-->A	24.012
C	289.046	C-->T	15.000
C	289.046	C-->G	40.006
5-Iodo-C	414.946	5-Iodo-C-->A	-101.888
5-Iodo-C	414.946	5-Iodo-C-->T	-110.900
5-Iodo-C	414.946	5-Iodo-C-->G	-85.894
G	329.052	G-->A	-15.994
G	329.052	G-->T	-25.006
G	329.052	G-->C	-40.006
G	329.052	G-->5-Iodo-C	85.894

[226] Mass spectra of bioagent-identifying amplicons were analyzed independently using a maximum-likelihood processor, such as is widely used in radar signal processing. This processor, referred to as GenX, first makes maximum likelihood estimates of the input to the mass spectrometer for each primer by running matched filters for each base composition aggregate on the input data. This includes the GenX response to a calibrant for each primer.

[227] The algorithm emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of *a priori* expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database is used to define the mass base count matched filters. The database contains the sequences of known bacterial bioagents and includes threat organisms as well as benign background organisms. The latter is used to estimate and subtract the spectral signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a running-sum estimate of the noise covariance. Background signal strengths are estimated and used along with the matched filters to form signatures which are then subtracted. The maximum likelihood process is applied

to this "cleaned up" data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

[228] The amplitudes of all base compositions of bioagent-identifying amplicons for each primer are calibrated and a final maximum likelihood amplitude estimate per organism is made based upon the multiple single primer estimates. Models of all system noise are factored into this two-stage maximum likelihood calculation. The processor reports the number of molecules of each base composition contained in the spectra. The quantity of amplification product corresponding to the appropriate primer set is reported as well as the quantities of primers remaining upon completion of the amplification reaction.

[229] Base count blurring can be carried out as follows. "Electronic PCR" can be conducted on nucleotide sequences of the desired bioagents to obtain the different expected base counts that could be obtained for each primer pair. See for example, ncbi.nlm.nih.gov/sutils/e-pcr/; Schuler, *Genome Res.* 7:541-50, 1997. In one illustrative embodiment, one or more spreadsheets, such as Microsoft Excel workbooks contain a plurality of worksheets. First in this example, there is a worksheet with a name similar to the workbook name; this worksheet contains the raw electronic PCR data. Second, there is a worksheet named "filtered bioagents base count" that contains bioagent name and base count; there is a separate record for each strain after removing sequences that are not identified with a genus and species and removing all sequences for bioagents with less than 10 strains. Third, there is a worksheet, "Sheet1" that contains the frequency of substitutions, insertions, or deletions for this primer pair. This data is generated by first creating a pivot table from the data in the "filtered bioagents base count" worksheet and then executing an Excel VBA macro. The macro creates a table of differences in base counts for bioagents of the same species, but different strains. One of ordinary skill in the art may understand additional pathways for obtaining similar table differences without undo experimentation.

[230] Application of an exemplary script, involves the user defining a threshold that specifies the fraction of the strains that are represented by the reference set of base counts for each bioagent. The reference set of base counts for each bioagent may contain as many different base counts as are needed to meet or exceed the threshold. The set of reference base counts is defined by taking the most abundant strain's base type composition and adding it to the reference set and then the next most abundant strain's base type composition is added until the threshold is met or exceeded. The current set of data was obtained using a threshold of 55%, which was obtained empirically.

[231] For each base count not included in the reference base count set for that bioagent, the script then proceeds to determine the manner in which the current base count differs from each of the base

counts in the reference set. This difference may be represented as a combination of substitutions, Si=Xi, and insertions, Ii=Yi, or deletions, Di=Zi. If there is more than one reference base count, then the reported difference is chosen using rules that aim to minimize the number of changes and, in instances with the same number of changes, minimize the number of insertions or deletions. Therefore, the primary rule is to identify the difference with the minimum sum (Xi+Yi) or (Xi+Zi), *e.g.*, one insertion rather than two substitutions. If there are two or more differences with the minimum sum, then the one that will be reported is the one that contains the most substitutions.

[232] Differences between a base count and a reference composition are categorized as one, two, or more substitutions, one, two, or more insertions, one, two, or more deletions, and combinations of substitutions and insertions or deletions. The different classes of nucleobase changes and their probabilities of occurrence have been delineated in U.S. Patent Application Publication No. 2004209260 (U.S. Application Serial No. 10/418,514) which is incorporated herein by reference in entirety.

Example 6: Codon Base Composition Analysis – Assay Development

[233] The information obtained by the codon analysis method of the present invention is base composition. While base composition is not as information-rich as sequence, it can have the same practical utility in many situations. The genetic code uses all 64 possible permutations of four different nucleotides in a sequence of three, where each amino acid can be assigned to as few as one and as many as six codons. Since base composition analysis can only identify unique combinations, without determining the order, one might think that it would not be useful in genetic analysis. However, many problems of genetic analysis start with information that constrains the problem. For example, if there is prior knowledge of the biological bounds of a particular genetic analysis, the base composition may provide all the necessary and useful information. If one starts with prior knowledge of the starting sequence, and is interested in identifying variants from it, the utility of base composition depends upon the codons used an the amino acids of interest.

[234] Analysis of the genetic code reveals three situations, illustrated in Tables 6A-C. In Table 6A, where the leucine codon CTA is comprised of three different nucleotides, each of the nine possible single mutations are always identifiable using base composition alone, and result in either a "silent" mutation, where the amino acid is not changed, or an unambiguous change to another specific amino acid. Irregardless, the resulting encoded amino acid is known, which is equivalent to the information obtained from sequencing. In Table 6B, where two of the three nucleotides of the original codon are the same, there is a loss of information from a base composition measurement compared to sequencing. In this case, three of the nine possible single mutations produce unambiguous amino acid choices, while the other six each

produce two indistinguishable options. For example, if starting with the phenylalanine codon TTC, then either one of the two Ts could change to A, and base composition analysis could not distinguish a first position change from a second position change. A first position change of T to A would encode an isoleucine and a second position change of T to A would encode a tyrosine. However no other options are possible and the value of the information would depend upon whether distinguishing an encoded isoleucine from a tyrosine was biologically important. In Table 6C, all three positions have the same nucleotide, and therefore the ambiguity in amino acid identity is increased to three possibilities. Out of 64 codon choices, 20 have three unique nucleotides (as in Table 6A), 40 have two of the same and one different nucleotide (as in Table 6B) and 4 have the same nucleotide in all three positions (as in Table 6C).

Table 6A: Wild Type Codon with Three Unique Nucleobases

Codon Description	Codon(s)	Codon Base Composition	Amino Acid Coded
WILD TYPE CODON	CTA	A1C1T1	Leu
Single Mutation	ATA	A2T1	Ile
Single Mutation	GTA	A1G1T1	Val
Single Mutation	TTA	A1T2	Leu
Single Mutation	CAA	A1C2	Gln
Single Mutation	CGA	A1G1C1	Arg
Single Mutation	CCA	A1C2	Pro
Single Mutation	CTG	G1C1T1	Leu
Single Mutation	CTC	C2T1	Leu
Single Mutation	CTT	C1T2	Leu

Table 6B: Wild Type Codon with Two Unique Nucleobases

Codon Description	Codon(s)	Codon Base Composition	Amino Acid Coded
WILD TYPE CODON	TTC	C1T2	Phe
Single Mutations	ATC, TAC	A1C1T1	Ile, Tyr
Single Mutations	GTC, TGC	G1C1T1	Val, Cys
Single Mutations	CTC, TCC	C2T1	Leu, Ser
Single Mutation	TTA	A1T2	Leu
Single Mutation	TTG	G1T2	Leu
Single Mutation	TTT	T3	Phe

Table 6C: Wild Type Codon Having Three of the Same Nucleobase

Codon Description	Codon(s)	Codon Base Composition	Amino Acid Coded
WILD TYPE CODON	TTT	T3	Phe
Single Mutations	ATT, TAT, TTA	A1T2	Ile, Tyr, Leu
Single Mutations	GTT, TGT, TTG	G1T2	Val, Cys, Leu
Single Mutations	CTT, TCT, TTC	C1T2	Leu, Ser, Phe

Example 7: Testing of HIV Strains for Drug Resistance by Codon Analysis

[235] Upon determination of an HIV positive test, drug resistance testing can be done by conventional sequencing methods to establish the nucleotide sequence of the major HIV strain infecting the patient followed by analysis of the codons most important in mediating drug resistance. The patient would then be monitored while on antiretroviral therapy by codon analysis methods according to the present invention for the appearance of emerging viral mutations. The advantages of monitoring by these methods for rapid codon analysis are: (i) it is much more sensitive to identifying low abundance mutations in a population, (ii) it can be done on a much lower viral titer and (iii) it is less expensive than sequencing. Based upon previous data on the sensitivity and dynamic range of the mass spectrometer, the ability is anticipated to identify a low-abundance mutation present in as little as 0.1% of the viral population in a 10 to 100-fold lower titer of virus than can be analyzed by sequencing.

[236] According to the 2005 update of drug resistance mutations in HIV there are 28 major and minor mutations in HIV reverse transcriptase and 21 major and minor mutations in HIV protease (Johnson, V. A. et al. 2005, Top. HIV Med. *13*, 51-57). A smaller subset of these mutations is currently used in making decisions regarding the drug regimen of choice for a particular virus strain. All 49 positions where mutations occur which are related to antiretroviral therapy were examined both with respect to the amino acid changes that mediate resistance and sequences present in isolates of the virus which included HIV-1 subtype B sequences from persons with well-characterized histories of antiretroviral treatment (Rhee, S. Y. et al. 2005, J. Infect. Dis. *192*, 456-65). Sequences were obtained from the Stanford HIV Reverse Transcriptase and Protease Sequence Database and included sequences from published studies and previously unpublished sequences generated at Stanford University from patients living in northern California (GenBank accession numbers AY796421–AY798497 and AY800656–AY802758). For the present example, sequences were aligned using an alignment editor and the relevant codons of the reverse transcriptase gene in 2,102 sequences were analyzed. An example of this analysis is illustrated in Table 7. Out of 2,102 HIV sequences, the majority have a wild type codon encoding methionine in position 41 of reverse transcriptase. Deviations from wild type observed were all changes that encode leucine, which is a thymidine nucleotide-associated mutation (TAM) associated with drug resistance.

Table 7: Mutations of Codon 41 in HIV Reverse Transcriptase

Codon	Total	Fraction	Amino Acid	Codon Base Composition	Phenotype	Distinguishable from Wild Type
ATG	1224	0.58	Met [M]	$A_1G_2T_3$	Wild type	-

TTG	429	0.20	Leu [L]	G_1T_2	TAM	yes
CTG	284	0.14	Leu [L]	$G_2C_1T_1$	TAM	yes
TTA	16	0.01	Leu [L]	A_1T_2	TAM	yes
CTA	7	0.00	Leu [L]	$A_1C_1T_1$	TAM	yes
CTT	1	0.00	Leu [L]	C_1T_2	TAM	yes
Subtotal	1961	0.93				
Mixtures	141	0.07				
Total	2102					

[237] From codon analysis of this dataset, several observations can be made. First, the wild type codon is the most frequently observed (in this case 58%). Second, the next most abundant codons are single point mutations from wild type, and encode a documented mutation which causes drug-resistance. Third, a significant fraction of the sequences (in this case 7%) contain ambiguities with respect to the correct nucleotide in the sequencing reaction (in Tables 6B and 6C, all sequences with more than one codon per base composition are grouped together). These may be derived from the presence of a mixed population of virus sequences in the original patient samples. Codon 41 of reverse transcriptase exemplifies the situation in Table 6A where the codon is comprised of three different nucleotides ATG, and all mutations are unambiguously identifiable with respect to the amino acid that they encode.

[238] The reverse transcriptase codon leucine 210 (Table 8) provides an example of the wild type codon base composition situation illustrated in Table 6B. Leucine is the most extreme example of degeneracy in the genetic code, and six different codons encode leucine. Nevertheless, mutation of the wild type codon TTG to TGG (tryptophan), which causes drug resistance, is unambiguously distinguishable from all six wild type leucine codons present, and no other codon in the dataset contains the composition base composition G_2T. Therefore, base composition analysis may be sufficient to distinguish a drug-resistant strain from the wild type strain, even in the absence of prior knowledge of the particular codon used to encode leucine.

Table 8: Mutations of Codon 210 in HIV Reverse Transcriptase

Codon	Total	Fraction	Amino Acid	Codon Base Composition	Phenotype	Distinguishable from Major or Minor Wild Type
TTG	1223	0.58	Leu [L]	G1T2	Major wild type	-
TGG	497	0.24	Trp [W]	G2T1	TAM	yes
TTA	168	0.08	Leu [L]	A1T2	Minor Wild type	-
CTG	17	0.01	Leu [L]	G1C1T1	Minor Wild type	-
TTC	9	0.00	Phe [F]	C1T2		
TCA	5	0.00	Ser [S]	A1C1T1		
CTC	4	0.00	Leu [L]	C2T1	Minor Wild type	-
CTT	2	0.00	Leu [L]	C1T2	Minor	-

					Wild type	
TCG	2	0.00	Ser [S]	G1C1T1		
CTA	1	0.00	Leu [L]	A1C1T1	Minor Wild type	-
GAG	1	0.00	Glu [E]	A1G2		
ATG	2	0.00	Met [M]	A1G1T1		
Subtotal	1931	0.92				
Mixtures	170	0.08				
Total	2101					

[239] Not all mutations are unambiguously identifiable by base composition in the absence of knowledge of the starting sequence. For example, reverse transcriptase lysine 65 can mutate to arginine, which mediates drug resistance. Lysine is an example of the situation in Fig. 2C, where one of the two codons that encode lysine have all three positions comprised of the same nucleotide (AAA), and the second lysine codon has two of the same and one different nucleotide (AAG). A mutation that encodes arginine AGA is not distinguishable from the minor wild type codon, which does not allow unambiguous assignment of drug resistance using base composition. However, if the starting virus wild type sequence is known (either AAA or AAG), then any mutation to arginine would be distinguishable.

Table 9: Mutations of Codon 65 in HIV Reverse Transcriptase

Codon	Total	Fraction	Amino Acid	Codon Base Composition	Phenotype	Distinguishable from Major or Minor Wild Type
AAA	1782	0.85	Lys [L]	A3	Wild type	
AAG	196	0.09	Lys [L]	A2G1	Minor wild type	
AGA	46	0.02	Arg [R]	A2G1	NNRTI*	Not distinguishable from minor wild type
AGG	1	0.00	Arg [R]	A1G1	NNRTI*	Yes
subtotal	2025					
Mixtures	77	0.04				
total	2102					

*NNRTI = nucleoside and nucleotide reverse transcriptase inhibitor resistant

[240] Thus, for most of the 49 mutations cited in the 2005 update of drug resistance mutation in HIV (Johnson, V. A. et al. 2005, Top. HIV Med. *13*, 51-57), base composition analysis is sufficient to determine whether a drug-resistant mutation is present, simply based upon the measured composition, the known biological constraints of the HIV virus and the rules of the genetic code. For a minority of the mutations, prior knowledge of the sequence of the particular virus strain is needed to distinguish mutations that mediate drug-resistance mutations from minor wild type codons variations.

[241] The advantages of monitoring codon base compositions include: sensitivity to identification of low abundance mutations in a population, it can be done on a much lower viral titer and is less expensive than sequencing. Based upon previous data on the sensitivity and dynamic range of electrospray mass spectrometry, we would anticipate the ability to identify a low-abundance mutation present in as little as 0.1% of the viral population in a 10 to 100-fold lower titer of virus than can be analyzed by sequencing. If a drug-resistant mutation should arise as a small fraction of the population in a patient that currently has a low viral titer, it could be valuable to change drugs sooner rather than later to keep the viral titer low. Lower virus titers mean there are fewer viruses available to mutate, and drug resistance would be suppressed.

[242] The present invention includes any combination of the various species and subgeneric groupings falling within the generic disclosure. This invention therefore includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[243] While in accordance with the patent statutes, description of the various embodiments and examples have been provided, the scope of the invention is not to be limited thereto or thereby. Modifications and alterations of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the present invention.

[244] Therefore, it will be appreciated that the scope of this invention is to be defined by the appended claims, rather than by the specific examples which have been presented by way of example.

[245] Each reference (including, but not limited to, journal articles, U.S. and non-U.S. patents, patent application publications, international patent application publications, gene bank accession numbers, internet web sites, and the like) cited in the present application is incorporated herein by reference in its entirety.

CLAIMS

What is claimed is:

1. An oligonucleotide primer 14 to 35 nucleobases in length comprising at least 70% sequence identity with SEQ ID NO: 47.

2. An oligonucleotide primer 14 to 35 nucleobases in length comprising at least 70% sequence identity with SEQ ID NO: 286.

3. A composition comprising the primer of claim 1.

4. The composition of claim 3 further comprising an oligonucleotide primer 14 to 35 nucleobases in length comprising at least 70% sequence identity with SEQ ID NO: 286.

5. The composition of claim 4 wherein either or both of said primers comprises at least one modified nucleobase.

6. The composition of claim 5 wherein said modified nucleobase is 5-propynyluracil or 5-propynylcytosine.

7. The composition of claim 4 wherein either or both of said primers comprises at least one universal nucleobase.

8. The composition of claim 7 wherein said universal nucleobase is inosine.

9. The composition of claim 4 wherein either or both of said primers further comprises a non-templated T residue on the 5'-end.

10. The composition of claim 4 wherein either or both of said primers comprises at least one non-template tag.

11. The composition of claim 4 wherein either or both of said primers comprises at least one molecular mass modifying tag.

12. A kit comprising the composition of claim 4.

13. The kit of claim 12 further comprising one or more primer pairs wherein each member of said one or more primer pairs is of a length of 14 to 35 nucleobases and has 70% to 100% sequence identity with the corresponding member from the group of primer pairs represented by SEQ ID NOs: 70:286, 165:286, 122:275, 100:336, and 61:324.

14. The kit of claim 12 further comprising at least one calibration polynucleotide.

15. The kit of claim 12 further comprising at least one anion exchange functional group linked to a magnetic bead.

16. A method for identification of a polyomavirus in a sample comprising:

amplifying nucleic acid from said polyomavirus using the composition of claim 4 to obtain an amplification product;

determining the molecular mass of said amplification product;

optionally, determining the base composition of said amplification product from said molecular mass; and

comparing said molecular mass or base composition with a plurality of molecular masses or base compositions of known polyomavirus identifying amplicons, wherein a match between said molecular mass or base composition and a member of said plurality of molecular masses or base compositions identifies said polyomavirus.

17. The method of claim 16 wherein said sample is a biological product.

18. A method of determining the presence or absence of a polyomavirus in a sample comprising:

amplifying nucleic acid from said sample using the composition of claim 4 to obtain an amplification product;

determining the molecular mass of said amplification product;

optionally, determining the base composition of said amplification product from said molecular mass; and

comparing said molecular mass or base composition of said amplification product with the known molecular masses or base compositions of one or more known polyomavirus identifying amplicons, wherein a match between said molecular mass or base composition of said amplification product and the molecular mass or base composition of one or more known polyomavirus identifying amplicons indicates the presence of said polyomavirus in said sample.

19. The method of claim 18 wherein said sample comprises a biological product.

20. A method for determination of the quantity of an unknown polyomavirus in a sample comprising:

contacting said sample with the composition of claim 4 and a known quantity of a calibration polynucleotide comprising a calibration sequence;

concurrently amplifying nucleic acid from said unknown polyomavirus and nucleic acid from said calibration polynucleotide in said sample with the composition of claim 4 to obtain a first amplification product comprising a polyomavirus identifying amplicon and a second amplification product comprising a calibration amplicon;

determining the molecular mass and abundance for said polyomavirus identifying amplicon and said calibration amplicon; and

distinguishing said polyomavirus identifying amplicon from said calibration amplicon based on molecular mass, wherein comparison of polyomavirus identifying amplicon abundance and calibration amplicon abundance indicates the quantity of polyomavirus in said sample.

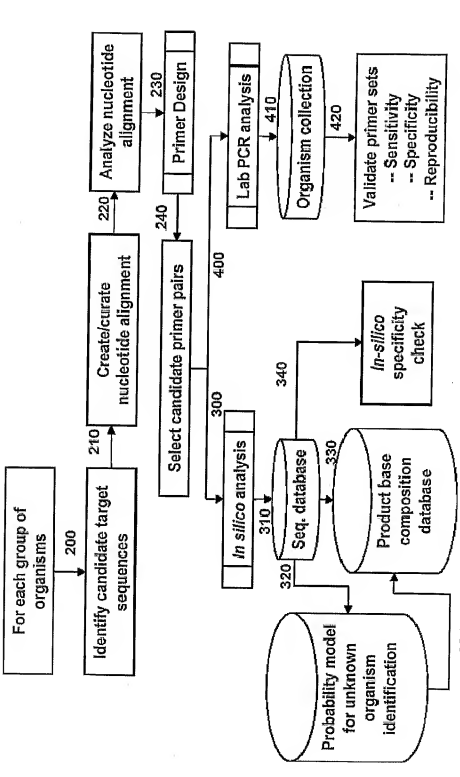
For each group of organisms
200
Identify candidate target sequences
210
Create/curate nucleotide alignment
220
Analyze nucleotide alignment
230
Primer Design
240
Select candidate primer pairs
400
Lab PCR analysis
410
Organism collection
420
Validate primer sets
-- Sensitivity
-- Specificity
-- Reproducibility
300
In silico analysis
310
320
Seq. database
330
Product base composition database
340
In-silico specificity check
325
Probability model for unknown organism identification

Figure 1

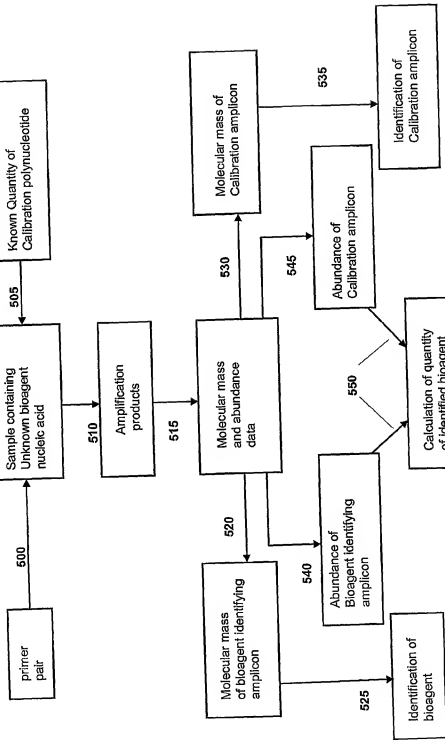
primer pair
500
Sample containing Unknown bioagent nucleic acid
505
Known Quantity of Calibration polynucleotide
510
Amplification products
515
Molecular mass and abundance data
520
Molecular mass of bioagent identifying amplicon
525
Identification of bioagent
530
Molecular mass of Calibration amplicon
535
Identification of Calibration amplicon
540
Abundance of Bioagent identifying amplicon
545
Abundance of Calibration amplicon
550
Calculation of quantity of identified bioagent

Figure 2

SEQUENCE LISTING

<110> Isis Pharmaceuticals, Inc.
Sampath, Rangarajan

<120> COMPOSITIONS FOR USE IN IDENTIFICATION OF ADVENTITIOUS VIRUSES

<130> DIBIS-0085WO

<150> 60/658,248
<151> 2005-03-03

<150> 60/705,631
<151> 2005-08-03

<150> 60/732,539
<151> 2005-11-01

<150> 60/740,617
<151> 2005-11-28

<160> 372

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 1
agcgtctagc catggcgtta gtat 24

<210> 2
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 2
ggaggctccg ttgtctgcat gta 23

<210> 3
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 3
taaccaagcc tttag 15

<210> 4
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 4
taagagtggg accatggata aacac 25

<210> 5

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 5
taagtaagtt caaccaagcc tttag 25

<210> 6
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 6
tacacacaat tcttggaag cacaggca 28

<210> 7
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 7
tacagatttc agcaagttca atcaagc 27

<210> 8
<211> 27
<212> DNA
<213> Artificial Sequence

<220>

<221> modified_base
<222> (22)...(22)
<223> I

<220>
<223> Primer

<400> 8
taccaacatc gcggaggcta ttgccca 27

<210> 9
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 9
tacccagtct acgtgtttgg cgactgtgt 29

<210> 10
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 10
taccccctgg gggtactttg a 21

<210> 11
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 11
taccctgct cgtgttacag g 21

<210> 12
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 12
tacggtcaca gcttcccaca t 21

<210> 13
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 13
tactaggagg ctgtaggcat aaattggt 28

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 14
tactcaccaa ctcctggccc 20

<210> 15
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 15
tactctcaca cggcctcata cagtac 26

<210> 16
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (10)...(10)
<223> I

<220>
<221> modified_base
<222> (19)...(19)

<223> I

<400> 16
tactgttatt caggatggtg atatggt 27

<210> 17
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (10)...(10)
<223> I

<400> 17
tactgttatt caggatggtg atatggt 27

<210> 18
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (19)...(19)
<223> I

<400> 18
tactgttatt caggatggtg atatggt 27

<210> 19
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 19
tactgttatt caggatggtg atatggt 27

<210> 20
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 20
tagaatgttt gctatgcaac c 21

<210> 21
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 21
tagaggcgga tgacaatggc g 21

<210> 22

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 22
tagatccaca agctttaggg tctg 24

<210> 23
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 23
tagatgatag tgacattgca tataaatatg ca 32

<210> 24
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (14)...(14)
<223> I

<220>
<221> modified_base
<222> (17)...(17)
<223> I

<220>
<221> modified_base
<222> (24)...(24)
<223> I

<400> 24
tagatgatag tgacattgca tataaatatg ca 32

<210> 25
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 25
tagcaggaag cactatgggc g 21

<210> 26
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 26
taggacccct gctcgtgtta cag 23

<210> 27
<211> 19

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 27
taggaccgtg gataaacac 19

<210> 28
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 28
taggatggcg atatggttga cacaggcttt gg 32

<210> 29
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (29)...(29)
<223> I

<400> 29
taggatggtg atatggttga tacaggcttt gg 32

<210> 30
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 30
taggatggtg atatggttga tacaggcttt gg 32

<210> 31
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (27)...(27)
<223> I

<220>
<221> modified_base
<222> (29)...(29)
<223> I

<400> 31
taggatggtg atatggttga tacaggcttt gg 32

<210> 32
<211> 32
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (24)...(24)
<223> I

<220>
<221> modified_base
<222> (27)...(27)
<223> I

<220>
<221> modified_base
<222> (29)...(29)
<223> I

<400> 32
taggatggtg atatggttga tacaggcttt gg 32

<210> 33
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 33
tagggagact ttttgctaaa atgac 25

<210> 34
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 34
tagggatttt tgacccatct ttttctca 28

<210> 35
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 35
taggggccca accccatttt cat 23

<210> 36
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 36
taggtagaat gtttgcaatg caacc 25

<210> 37
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 37
taggtcgcgt aatttgggta aggtcatc 28

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 38
tagtctcaaa gagaaagaga tcaaacaaga 30

<210> 39
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 39
tagtgcaatc catctagcag ctgt 24

<210> 40
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 40
tataatccac ctatcccagt aggagaaat 29

<210> 41
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 41
tatagggtca tgaatcaaga acccgg 26

<210> 42
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (24)...(24)
<223> I

<400> 42
tatagggtca tgaatcaaga acccgg 26

<210> 43
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 43
tatatatcac gagatctgca gtttatgagt aa 32

<210> 44
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 44
tatatcacgt gatctgcagt ttatgag 27

<210> 45
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 45
tatcagtgca atccatctag cagctgt 27

<210> 46
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 46
tatcgctgga tgtgtctgcg g 21

<210> 47
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 47
tatcttatca tgtctgggtc cccaggaag 29

<210> 48
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (18)...(18)
<223> I

<400> 48
tatggtgcag tgggcatttg ataatg 26

<210> 49
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 49
tatggtgcag tgggcatttg ataatg 26

<210> 50
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 50
tatgtcaaaa gctgtggaca atgat 25

<210> 51
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 51
tatgtttgct atgcaacc 18

<210> 52
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 52
tcaaccacat aagaggtggg tgttc 25

<210> 53
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 53
tcaacctcca atcactcacc aac 23

<210> 54
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 54
tcaagagatc ttcagtttat 20

<210> 55
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 55

tcaagagatc ttcagtttat gagtaa 26

<210> 56
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 56
tcaagatggg gaatgagatt gcccct 26

<210> 57
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 57
tcaatgtatc ttatcatgtc tgggtcccc 29

<210> 58
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 58
tcaccaaggt gcctctaaaa cga 23

<210> 59
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 59
tcaccacgca atgcttccct atctt 25

<210> 60
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 60
tcacctggga ccccatcgat ggac 24

<210> 61
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 61
tcacctttgc aaaggggccc 20

<210> 62
<211> 26
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<400> 62
tcacgagatc tgcagtttat gagtaa 26

<210> 63
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 63
tcagagcatc agatcacctg ggacc 25

<210> 64
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 64
tcaggatggt ttttggtaga ggctatagt 29

<210> 65
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 65
tcaggcaaga gtcactgcta tcgaga 26

<210> 66
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 66
tcagtgaata tgactagcca ggtgct 26

<210> 67
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 67
tcagtgtagg tagaatgttt gcaatgc 27

<210> 68
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 68
tcagttgatg ggcctacctc atttctt 27

<210> 69
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 69
tcatctacaa catgatggga aagagagaga 30

<210> 70
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 70
tcatgtctgg gtcccctgga ag 22

<210> 71
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 71
tccaaccaga aatttgagaa cattgcaga 29

<210> 72
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 72
tccacctgtg gttattgaac ctgt 24

<210> 73
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 73
tccaggtact tcacccccaa acctg 25

<210> 74
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 74
tccatacttc gccttcacca ttccc 25

<210> 75
<211> 25

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 75
tcccacatct gcaatcggag acagg 25

<210> 76
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 76
tccccaatct ccaatcactc accaa 25

<210> 77
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 77
tccccatcag acacctcagg tgga 24

<210> 78
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 78
tccccggtga ttggtatcct cc 22

<210> 79
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 79
tccctctaca gtagcaacgg atgcaa 26

<210> 80
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 80
tccgatgcac attcacggtt ggtat 25

<210> 81
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 81
tccgccttcg atgacagacg g 21

<210> 82
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 82
tccggccctg gatgacaa 18

<210> 83
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 83
tccgggctgt caatatgcta aaacg 25

<210> 84
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 84
tccgtcgcaa gttggtgatg agta 24

<210> 85
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 85
tcctaagagt gggaccatgg ataaa 25

<210> 86
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 86
tcctaagagt gggaccatgg ataaacac 28

<210> 87
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 87
tcctagaagt agaggcagca tcca 24

<210> 88

<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 88
tcctcccaca gcaaacatgt g 21

<210> 89
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 89
tcctcgggaa attggcattg cgat 24

<210> 90
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 90
tcctctcact catcagagca cccc 24

<210> 91
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 91
tcctgcccac ctacaacaac ca 22

<210> 92
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 92
tccttcacgg aggctatgac taggta 26

<210> 93
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 93
tcgaaaaacc tccaggcaag agtcac 26

<210> 94
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 94
tcgcaagact gctagccgag ta 22

<210> 95
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 95
tcgcagccat cttggcaaca tcctc 25

<210> 96
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 96
tcgcagtccc caatctccaa tcact 25

<210> 97
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 97
tcggcgcaag tcccagaacc 20

<210> 98
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 98
tcgggatgta atggctggtt 20

<210> 99
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 99
tcgttatcgg ctcagctcta cagttc 26

<210> 100
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 100
tctacagtag caacggatgc aa 22

<210> 101
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 101
tctagccatg gcgttagtat gagtgt 26

<210> 102
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 102
tctcccgcac cctcaacaag c 21

<210> 103
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 103
tcttttccag accccggact cc 22

<210> 104
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 104
tgaacatacc aatgcagtt 19

<210> 105
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 105
tgaactaacg gacagtggat atggc 25

<210> 106
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 106
tgaatggttc cgccagtgca c 21

<210> 107
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 107
tgacacaggc acccacggaa ta 22

<210> 108
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 108
tgacgaacca cagcgtcaca 20

<210> 109
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 109
tgacgtagcc aaggaggcgt t 21

<210> 110
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 110
tgagaccatc ataagtagca agatgt 26

<210> 111
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 111
tgagcagcag gaagcactat gg 22

<210> 112
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 112
tgaggcctat agcagctata gcctc 25

<210> 113
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 113
tgatgaaata tcgtggatgg aagc 24

<210> 114
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 114
tgatgagata tcgtggatgg aagc 24

<210> 115
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 115
tgattgaccc ttttcagttg ggcctt 26

<210> 116
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 116
tgattgatgt tgaaaccacc gaaga 25

<210> 117
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 117
tgcagtcccc aatctccaat cact 24

<210> 118
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 118
tgccaatcac tcatacaacc cccaa 25

<210> 119
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 119
tgccacgacc ggcaagacca acat 24

<210> 120
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 120
tgccagatga tcctataatt gttgagga 28

<210> 121
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 121
tgccagggag tagtagaagc aatgaa 26

<210> 122
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 122
tgccattcat aggctgccca tc 22

<210> 123
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 123
tgcctacatt ctatcttgcg ttacatgg 28

<210> 124
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 124
tgcctagtgt gaagatgggg aatgagat 28

<210> 125
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 125
tgcctgaatt ggagatatga gaccat 26

<210> 126
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 126

tgccttaatt ggagatatga gac 23

<210> 127
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 127
tgccttaatt ggagatatga gaccat 26

<210> 128
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 128
tgcgttaact ggaccaatga gaactttcc 29

<210> 129
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 129
tgctcctctc aaaaccatgg gacac 25

<210> 130
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 130
tgctgaattc gtaacattga tga 23

<210> 131
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 131
tggaaatcct ttttctcaag gacgtggt 28

<210> 132
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 132
tggacattca tctctatcag tgc 23

<210> 133
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<400> 133
tggaccatca ttggaaggga cg 22

<210> 134
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 134
tggaggctcc gttgtctgca tgta 24

<210> 135
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 135
tggatagagg agaatgaatg gatggaagac 30

<210> 136
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 136
tggatgtgtc tgcggcgtt 19

<210> 137
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 137
tggcccgtg tctggtagc 19

<210> 138
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 138
tggctaggtg tcacgtagaa aggactac 28

<210> 139
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 139
tggctgaact tttgaaagtg agcgg 25

<210> 140
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 140
tggctgctag gctgtgctgc caact 25

<210> 141
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 141
tgggatgatc tactgagact gtgtga 26

<210> 142
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 142
tgggggtatt ttgacttcaa ccgattccac 30

<210> 143
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (22)...(22)
<223> I

<220>
<221> modified_base
<222> (24)...(24)
<223> I

<400> 143
tgggggtatt ttgacttcaa ccgattccac 30

<210> 144
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 144
tgggtcctgc ccacctacaa caacca 26

<210> 145
<211> 21
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (4)...(4)
<223> I

<400> 145
tggggtggtg agagtttggg a 21

<210> 146
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 146
tggtactcct caactgctgt cc 22

<210> 147
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 147
tggtgccgac ggagtggg 18

<210> 148
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 148
tggtgccgat ggagtggg 18

<210> 149
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 149
tggtgtggtg agagtttggg a 21

<210> 150
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 150
tggtgttgtg agggtctggg a 21

<210> 151
<211> 26
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<400> 151
tgtaaatcta gtggctctcc tcccac 26

<210> 152
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 152
tgtacaatct atgtaggaga tccttactgt cc 32

<210> 153
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 153
tgtaggtaga atgtttgcaa tgc 23

<210> 154
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 154
tgtaggtgat cccttcaatc ctcc 24

<210> 155
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 155
tgtcacactt gcatttacaa catgatgg 28

<210> 156
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 156
tgtcccgact gtggcacttg a 21

<210> 157
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 157
tgtctatctt cctgactc 18

<210> 158
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (9)...(9)
<223> I

<220>
<221> modified_base
<222> (12)...(12)
<223> I

<220>
<221> modified_base
<222> (18)...(18)
<223> I

<400> 158
tgtctctctg gcgatgggtg gcc 23

<210> 159
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 159
tgtctctctg gagatgggtg gcc 23

<210> 160
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 160
tgtctttcag gagatggatg gcc 23

<210> 161
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (12)...(12)
<223> I

<220>
<221> modified_base
<222> (15)...(15)
<223> I

<400> 161
tgtctttcag gagatggatg gcc 23

<210> 162
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 162
tgtgaagcgc acaattgtta agac 24

<210> 163
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 163
tgtgaatatc aagcaggaca taacaaggta gg 32

<210> 164
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 164
tgtgcatctt actcactaaa ggagaa 26

<210> 165
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 165
tgtgccctca aaaaccctaa cctc 24

<210> 166
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 166
tgtgccctca aaaaccctga cctc 24

<210> 167
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 167
tgtggacaat gatctccatt gctgcaat 28

<210> 168
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 168
tgtgtcatct gcgagtgtgg 20

<210> 169
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 169
tgttaatgtc tatcttcctg actc 24

<210> 170
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 170
tgtttgcaca gaggctaaat ga 22

<210> 171
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 171
ttacaaagag ggccaacgcc attctcatc 29

<210> 172
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 172
ttatgtcctt ggggtcacct actgg 25

<210> 173
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 173
ttcaatcaag catttcggta tgaaac 26

<210> 174
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 174

ttcacgcaga aagcgtctag cca 23

<210> 175
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 175
ttcacgcaga aagcgtctag ccat 24

<210> 176
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 176
ttcagatgtc tgtgtggcgg ccta 24

<210> 177
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (19)...(19)
<223> I

<400> 177
ttcagatgtc tgtgtggcgg ccta 24

<210> 178
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (12)...(12)
<223> I

<220>
<221> modified_base
<222> (19)...(19)
<223> I

<400> 178
ttcagatgtc tgtgtggcgg ccta 24

<210> 179
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 179
ttgcaaagag gcccaacccc attctcat 28

<210> 180
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (24)...(24)
<223> I

<400> 180
ttgcaaagag gcccaacccc attctcat 28

<210> 181
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 181
ttgcgtcaat ggcttatatc aaagg 25

<210> 182
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 182
ttggatcagc cactgatgaa agatc 25

<210> 183
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 183
ttggtccatg actccgacct tgaa 24

<210> 184
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 184
tttcgttgtg gcaaggtagg acac 24

<210> 185
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 185
tttggacacc caatggt 17

<210> 186
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 186
taaggaacgg cgacaggatg cgc 23

<210> 187
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 187
taatagactt tcccctctat attctaattc 30

<210> 188
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 188
taataggctt tcccctctat attctaattc 30

<210> 189
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 189
taatccagaa gaaccaacaa gaagatgagg 30

<210> 190
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 190
taattgcatt agtctcact 19

<210> 191
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 191
tacagaatag gctttg 16

<210> 192
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 192
tacatattca tccgtgctta caaccttaga 30

<210> 193
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 193
taccatgtcc gaacctgtat ctgt 24

<210> 194
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 194
tacggagagt gacgggtttc c 21

<210> 195
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 195
tacgggacgt aaacaaagga cgtcc 25

<210> 196
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 196
tacggggcca tgccgtgttg 20

<210> 197
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 197
tactcaccgg ttccgcagac cactat 26

<210> 198
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 198

tacttcgtct cgtcacgact tctgtt 26

<210> 199
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 199
tacttctaaca gcatccat 18

<210> 200
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 200
tacttcttgtt gtcacagcta tagcttgatt 30

<210> 201
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 201
tactgcataa tagactttcc cctctatatt ctaa 34

<210> 202
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 202
tactgcataa taggctttcc cctctaaatt ctaa 34

<210> 203
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 203
tacttgggat tgtttgtgtg agacgg 26

<210> 204
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 204
tagatttcat ttagcctctg tgcaaa 26

<210> 205
<211> 26
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<400> 205
tagcaaattg ctttgctgtt gcacta 26

<210> 206
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 206
tagctccgag ccacatgaac c 21

<210> 207
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 207
tagctgaaat tgctgctgga gagg 24

<210> 208
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 208
taggaatatg ataaaacgcc gcagacac 28

<210> 209
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 209
tagggagatt gggagctccc gtatt 25

<210> 210
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 210
tagggtcact tggaggattg aaagg 25

<210> 211
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 211
tagtattttg tcctgccacg catttaaacg 30

<210> 212
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (20)...(20)
<223> I

<400> 212
tagtattttg tcctgccacg catttaaacg 30

<210> 213
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (19)...(20)
<223> I

<400> 213
tagtattttg tcctgccacg catttaaacg 30

<210> 214
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (18)...(20)
<223> I

<400> 214
tagtattttg tcctgccacg catttaaacg 30

<210> 215
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 215
tagtcgttgg tcgttgtta gggaa 25

<210> 216
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 216

tatatgataa aacgccgcag atac 24

<210> 217
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 217
tatccccgaa agcccagata tatac 25

<210> 218
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 218
tatttgcctg ccaattgctt tttaaaaa 28

<210> 219
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (12)...(13)
<223> I

<220>
<221> modified_base
<222> (20)...(20)
<223> I

<400> 219
tatttgcctg ccaattgctt tttaaaaa 28

<210> 220
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 220
tcaaacccag aggtgcctgt aaa 23

<210> 221
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 221
tcaacactgg tgttgtccca 20

<210> 222
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 222
tcaatctctc ccttaaccat cccatt 26

<210> 223
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 223
tcactcgttt agacggcgga atatc 25

<210> 224
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 224
tcagcaaatt gttctgctgc tgcacta 27

<210> 225
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 225
tcagctatca atttctctgc caatatttg 29

<210> 226
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 226
tcagctatca ttttctcagc caagatttg 29

<210> 227
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 227
tcaggagctt tcgtgtccac ctt 23

<210> 228
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 228

tcaggcacag tcgggaggtc 19

<210> 229
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 229
tcaggctgcc acaccagatg tc 22

<210> 230
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 230
tcaggggtcc ttctata 17

<210> 231
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 231
tcaggggtgc tttgtgcttc 20

<210> 232
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 232
tcaggtggtg tctgccctcg tt 22

<210> 233
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 233
tcattgtggt gggtaggtcg tc 22

<210> 234
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 234
tcatttagcc tctgtgcaaa 20

<210> 235
<211> 25
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<400> 235
tccacatgaa ccaaatggct ctgct 25

<210> 236
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 236
tccagaagaa ccaacaagaa gatgaggc 28

<210> 237
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (11)...(11)
<223> I

<400> 237
tccagaccca cccaagaatg agaacacaag ata 33

<210> 238
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 238
tccagaccct gcaaaaaatg agaacac 27

<210> 239
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 239
tccatatacg aacacaccgg cgac 24

<210> 240
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 240
tccatcgatg gggtcccagg t 21

<210> 241
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<400> 241
tccatgggga atcgcaatgc caat 24

<210> 242
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 242
tccatttgga atcgcaatgc caat 24

<210> 243
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 243
tccgatgtgg cttatttggc tg 22

<210> 244
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 244
tccgtggatt gttctgtagc agtcttc 27

<210> 245
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 245
tcctattcct ccccttcttt taaaattcat gc 32

<210> 246
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 246
tcctcggtag tcctttctac gtg 23

<210> 247
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 247
tcctgcttga tggcctgtaa gtct 24

<210> 248
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 248
tcctttgcca tcccattgtc 20

<210> 249
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 249
tcgacacatt ggaggagcat gatgtta 27

<210> 250
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 250
tcgatcaggc tcaacatagc cctctt 26

<210> 251
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 251
tcgcaagcac cctatcaggc ag 22

<210> 252
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 252
tcgccatgtc tctctacctg cg 22

<210> 253
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 253
tcggcagcca tttgcaaata atcaggatat tt 32

<210> 254
<211> 20

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 254
tcgggagggc aatgagggtc 20

<210> 255
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 255
tcgggcccac tatgacgtgt aca 23

<210> 256
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 256
tcgggtactg gtgcacacgc 20

<210> 257
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 257
tcgtttgtag ggaacattgg tgaggaag 28

<210> 258
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 258
tcgtttgtag ggaacattgg tgaggaag 28

<210> 259
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 259
tctaaacgca catcccatg aattt 25

<210> 260
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 260
tctaaatgag gacctgacct gtg 23

<210> 261
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 261
tctaatattg tgtttatcca 20

<210> 262
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 262
tctatgagga cattttccca atctcc 26

<210> 263
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 263
tctcccatca tagtatatcc ttttgct 27

<210> 264
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 264
tctctgggtg gggaaggagg ggag 24

<210> 265
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 265
tctgcaacca tttgaaaata atctggatat tt 32

<210> 266
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 266
tctgcaacca tttgaaaata atctggatat tt 32

<210> 267

<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 267
tctgcaacca tttgcaaata atctggatat tt 32

<210> 268
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 268
tctgcaacca tttgcaaata atctggatat ttgca 35

<210> 269
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (33)...(33)
<223> I

<400> 269
tctgcaacca tttgcaaata atctggatat ttgca 35

<210> 270
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (15)...(15)
<223> I

<400> 270
tctgcaacca tttgcaaata atctggatat tt 32

<210> 271
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (15)...(15)
<223> I

<220>
<221> modified_base
<222> (33)...(33)
<223> I

<400> 271
tctgcaacca tttgcaaata atctggatat ttgca 35

<210> 272
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (14)...(15)
<223> I

<400> 272
tctgcaacca tttgcaaata atctggatat tt 32

<210> 273
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (14)...(15)
<223> I

<220>
<221> modified_base
<222> (33)...(33)
<223> I

<400> 273
tctgcaacca tttgcaaata atctggatat ttgca 35

<210> 274
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (14)...(15)
<223> I

<220>
<221> modified_base
<222> (17)...(17)
<223> I

<400> 274
tctgcaacca tttgcaaata atctggatat tt 32

<210> 275
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 275
tctggaaccc agcagtgga 19

<210> 276
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 276
tcttcatcag agctaggaga ttctgttgt 29

<210> 277
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 277
tcttgctttg acatgttgtg gtgga 25

<210> 278
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 278
tctttcccct ctgtattcta a 21

<210> 279
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 279
tgaaaacaca agatactttg gaacatacac aggaggt 37

<210> 280
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 280
tgaaccaatt gcattagtct cact 24

<210> 281
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 281
tgaagaggaa tatgataaaa cgccgcaga 29

<210> 282
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 282
tgaagggaaa gttctcattg gtccagtt 28

<210> 283
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 283
tgaagtcatc cagtatagtg tttatcca 28

<210> 284
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 284
tgaatgtacc ccatgaggtc ggc 23

<210> 285
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (18)...(18)
<223> I

<220>
<221> modified_base
<222> (27)...(27)
<223> I

<220>
<221> modified_base
<222> (34)...(34)
<223> I

<400> 285
tgagaacaca agatactttg gaaactttca caggtgg 37

<210> 286
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 286
tgagagtgga tgggcagcct atg 23

<210> 287
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 287
tgaggatgtg tccaagatgg ctgcg 25

<210> 288
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 288
tgaggctgga tctatccacg caaa 24

<210> 289
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 289
tgaggggagt cgagggataa ggaac 25

<210> 290
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 290
tgaggggat aatggcatag gaggaat 27

<210> 291
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 291
tgagtgattg gcggggtaag gac 23

<210> 292
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 292
tgataaaacg ccgcagacac atc 23

<210> 293
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 293
tgattgtcgc cttgaaccat tgc 23

<210> 294
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 294
tgcacacaac ggacacacaa a 21

<210> 295
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 295
tgcactagaa aacttcatct gggttgcagt atc 33

<210> 296
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 296
tgcactagag aatttcatct gggttgcagt atc 33

<210> 297
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 297
tgcagcgtag gaaggtgtgg 20

<210> 298
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 298
tgcatagcaa tctcaggagg tgc 23

<210> 299
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 299
tgcatgaata agggtctgaa gccca 25

<210> 300
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 300
tgcatgttga tgttgtccag catgat 26

<210> 301
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 301
tgccgaggcc tatgtggtcg acatt 25

<210> 302
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 302
tgcctgtgct tccaaggaat tgtgtgtaat a 31

<210> 303
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 303
tgcgggaaag gcgggagttg aagat 25

<210> 304
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 304
tgctaacaat ttctcagc 18

<210> 305
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 305
tgctccagga ggtgcaaatc aaaga 25

<210> 306
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 306

tgctcccccc agaaaattga gagaagtc 28

<210> 307
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 307
tgctgatctt ggctgtccac ctc 23

<210> 308
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 308
tgctggcatg aaccaaactg att 23

<210> 309
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 309
tgcttccatc cacgatatct catc 24

<210> 310
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 310
tggaccatac aggcaacccg acaa 24

<210> 311
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 311
tggcaattcc ggtgtactca cc 22

<210> 312
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 312
tggcagtgga atcggttgaa gtcaaa 26

<210> 313
<211> 21
<212> DNA

<213> Artificial Sequence

<220>
<223> Primer

<400> 313
tggcatggtg ccaaactgat t 21

<210> 314
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (7)...(7)
<223> I

<400> 314
tggcatggtg ccaaactgat t 21

<210> 315
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 315
tggcttgatt gtctccttga accattgc 28

<210> 316
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 316
tgggaaagct ggtagaggta catgc 25

<210> 317
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 317
tgggactctg gttcatgtat tgg 23

<210> 318
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 318
tgggcaatga gggtcact 18

<210> 319
<211> 21

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 319
tgggcctctc tgcaaaggag a 21

<210> 320
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 320
tgggctgata tggaatgtgc cctatctg 28

<210> 321
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 321
tggggagatg accatgatgg tgc 23

<210> 322
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 322
tggggctcat ggtcattgtc atc 23

<210> 323
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 323
tgggggacct aattgctact gtatctga 28

<210> 324
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 324
tgggtccctg atccaactag aaatgaaaa 29

<210> 325
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 325
tggtaatcaa aatactgcgg gccaa 25

<210> 326
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 326
tggtacttctt gatgtcacag ctatagcttg att 33

<210> 327
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 327
tggtcaaagc tggcatgaac caaa 24

<210> 328
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (7)...(7)
<223> I

<220>
<221> modified_base
<222> (11)...(11)
<223> I

<400> 328
tggtcaaagc tggcatgaac caaa 24

<210> 329
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 329
tggtctggaa aagacagggt tggg 24

<210> 330
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 330
tggtcttctg gggcttgttc catc 24

<210> 331
<211> 25

<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 331
tggtgtttgg agtggctatt ggcag 25

<210> 332
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 332
tgtaaccagc agaataggct ttg 23

<210> 333
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 333
tgtaactgtt aaaactgagg ttgttggagt g 31

<210> 334
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 334
tgtacggcag tgtttgcggt at 22

<210> 335
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 335
tgtagaggtg gttgttgtag gtggg 25

<210> 336
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 336
tgtatctgaa gctgctgctg c 21

<210> 337
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 337
tgtcatatcc cctattcctc cccttctt 28

<210> 338
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 338
tgtcatcttg ccctcctccc acca 24

<210> 339
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 339
tgtcattttg ccctcctccc acca 24

<210> 340
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 340
tgtccataat ttctggcaat aaccttctat 30

<210> 341
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 341
tgtctaaacg cacatcccca tgaattt 27

<210> 342
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 342
tgtctagact ctgtggtatt gtgagga 27

<210> 343
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 343
tgtgtctgta aagactgaag ttgttgga 28

<210> 344

<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 344
tgtgtttttcc caagctcttc c 21

<210> 345
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 345
tgttccactt cccaagttta catcttcctc 30

<210> 346
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 346
tgttccgcag actactatgg ctctc 25

<210> 347
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 347
tgttgaagtc aaaatacccc caggggggt 28

<210> 348
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 348
tgttgaggac ggctgctaat tgttg 25

<210> 349
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 349
tgttgtgata acaaagtccc aatcatcgtt c 31

<210> 350
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

"<223> Primer"

<400> 350
tgttgtgcac ttttggagaa tattt 25

<210> 351
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (16)...(16)
<223> I

<400> 351
tgttgttgat gagtctctgc cagtc 25

<210> 352
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 352
tgttgttgat gagtctctgc cagtc 25

<210> 353
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 353
tgtttatcca cggtcccact c 21

<210> 354
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 354
tgtttatcca tggtcccact c 21

<210> 355
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (11)...(11)
<223> I

<400> 355
tgttttccca tgccctccc 19

<210> 356

<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 356
ttaggtcctg cacagccgca taatg 25

<210> 357
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 357
ttctcagcta acaatttctc agc 23

<210> 358
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 358
ttctgcaatt acttgacacg acctcat 27

<210> 359
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 359
ttgcagagat ggagatcatt gtcca 25

<210> 360
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 360
ttgcttggtt atcaccctga acca 24

<210> 361
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 361
ttgcttttta aaaatgcagc tgcatt 26

<210> 362
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (20)...(21)
<223> I

<400> 362
ttgctttttta aaaatgcagc tgcatt 26

<210> 363
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (20)...(21)
<223> I

<400> 363
ttgctttttta aaaatgcagc tgcatt 26

<210> 364
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> modified_base
<222> (18)...(21)
<223> I

<400> 364
ttgctttttta aaaatgcagc tgcatt 26

<210> 365
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 365
ttggcggggt gaggaccttg aggg 24

<210> 366
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 366
ttgtcaatat catccagttg gc 22

<210> 367
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 367
ttgtgcacca tgcagttcat c 21

<210> 368
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 368
tttctgctcg tttataatgt ctacacat 28

<210> 369
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 369
tttgagttgt gggcagtccc ttt 23

<210> 370
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 370
tttggtcctt gtcttatgtc cagaatg 27

<210> 371
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 371
ttttggcgaa tattttgttt gg 22

<210> 372
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 372
tgttccgcag accactatgg ctctc 25